

-599-

Query: 130 TPTYTNGSVLIVNRLAKGLTIKSYEDLLQPSLKGKIAFPADPNTSSSAFSLQITNILLAKG 189
 T ++ S+L+VN LA + I+ YEDLL P LKGKIA ADE+ SSSAP L N+L A G
 Sbjct: 122 TRFSAIPSLIVNRLNAGNLIKIBGYEDLLNPELKGKIAADPSASSAPSHLVNMLIYAG 181

5 Query: 190 GYTNPKANMYVKKLQHNDAIKSSSSSEVYQSVARGKMIVGLTYEDPSVNLQKSGANVSI 249
 K W+YV+KL N++ S SS VY+ VA+G+ VGLTYE+P ++ SG+ V +
 Sbjct: 182 KGDPEKGDVYVQKLCANLDGKILGSSSAVYKGVADGETYVGLTYEKPGLSYMSSGSPVK 241

10 Query: 250 VYPTBGTVPFSSVAIIKGAISMKEAKLFINFMLSLDVQNAFQGSTSNRPIKDAQTSNG 309
 +Y EG + P V IIK +++ AK FI++ +SLD QN + S R IR DA ++
 Sbjct: 242 TYMKGVLSKFDGVYIIKGGKILENAKKFIYCVSLDAQNMVLEKLSRRSIRSDAVVTDM 301

Query: 310 MKALKDIATLKEDYRYVYKHGQILKTYNRI 340
 +K + +I ++ ++ V + + + L + I
 15 Sbjct: 302 VKPMSEIYSITNDVVEESQKWLQKFKDI 332

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1617> which encodes the amino acid sequence <SEQ ID 1618>. Analysis of this protein sequence reveals the following:

Possible site: 33

20 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.16 Transmembrane 9 - 25 (4 - 33)

----- Final Results -----
 25 bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

30 >GP:AAB95371 GB:U57349 periplasmic-iron-binding protein Btc
 [Brachyspira hyodysenteriae]
 Identities = 115/324 (35%), Positives = 177/324 (54%), Gaps = 8/324 (2%)

Query: 15 VIIILAIIVNMYIP-----SSSHKDSAKELVILTPNSQTILVOTIPAFSEKY-GVKVRL 68
 +++I + +++++IP S S S LVI P+ + + F+K G+ V +
 35 Sbjct: 4 IVLIFTSLILSVFIPYFSCSSSSGAQSGNSLVYCPHLEFINPLVDDPKANPQINVDI 63

Query: 69 IQGGTGLIDQL-GRKKDPINADIPFGNVTQFESHKDLFESVUSPQNTVISDYQLPSH 127
 I GTG+L+ ++ KD PL DI +GG + + DLPEY S + Y+
 40 Sbjct: 64 IAACTGELLKRVSEKINPLG-DILAGGTISMAKPKIDLPESTYSTNRENIARIYNTG 122

Query: 128 RATPYTNGSVLIVNRLAKGLTIKSYEDLLQPALKGKIAFPADPNTSSSAFSLQITNILLAKG 187
 T T S+L+VN LA + I YEDLL P LKGKIAFADP++SSS+P L N+L A
 45 Sbjct: 123 ALTRCTAVPSILMVNRLNAGDKIBGYEDLLNPELKGKIAFADPSSASSAPSHLVNMLYA 182

Query: 188 KGGYTNADAVMYNKRLLVNMNSIRATSSSEVYQSVARGKMIVGLTYEDPCINLQKSGANV 247
 G N Y+ +L N++ + SS VY+ VA+G+ VGLT+ B+ N +G+ V
 Sbjct: 183 IKGSDPEKGDVYVSKLCANLDGKILGSSSAVYKGVADGETYVGLTYEKPGLSYMSSGSPVK 242

50 Query: 248 SLVYKPGSTVPFSSVAIIKGAISMKEAKLFINFMLSRDVQNAFQGSTSNRPIKDAQTS 307
 +VY KEG + P + IIK+A N+ AK F++ S D Q +R +R D S
 Sbjct: 243 KLVYMKGVYIIKPGSIYIIKNAKLENAKKFVDYATSYDAQKTYITDKLNRREBVGULPFS 302

Query: 308 HMKALETTIATLKEDYRYVYKHKK 331
 55 +++++TI + +D A V ++K+
 Sbjct: 303 ALQSVDTINVTIDSAVDVQNKQ 326

An alignment of the GAS and GBS proteins is shown below:

Identities = 257/345 (74%), Positives = 295/345 (85%), Gaps = 1/345 (0%)

60 Query: 1 MKEKQSKRLIYILLVNSIIIPGVFTYSIQPSKLPKPRVILGPSNQALINGTIPAFEE 60
 +K K+ L++L+++ + +V Y S SK KLVIL+PNSQ ILGVTIPAFEE 60
 Sbjct: 2 LKLRKRLISPLVILVILVNVNMYIPSS-SKKDSAKELVILTPNSQTILQITIPAFEE 60

-600-

Query: 61 KYGIKVKLIQGGTGLIDRLSKGKGLKADIFFGGNTYQFSSHKALPESVYGRNVHTVIF 120
 KYG+KV+LIQGGTGLID+L ++ K L ADIFFGGNTYQFSSHK LPSVYS V TVI
 Sbjct: 61 KYGVKVKLIQGGTGLIDQLGRKDKPLNADIFFGGNTYQFSSHKDLPSVSPQVSTVIS 120

5 Query: 121 DYIHPSITATPYTTINGSVLIVNNELARGLTIKSYEDLIQPSLKGKIAFADPNTSSSAFQS 180
 DY PS ATPYTTINGSVLIVNNELA+GL I SYEDLQP+LKGKIAFADPN+SSSAFQS
 Sbjct: 121 DYQLPSHRTATPYTTINGSVLIVNNELARGSLHITSYEDLIQPALKGKIAFADPNSSSSAFQS 180

10 Query: 181 LTNILLAKGSGYTNPKAWNYVKLIQHNIMAKSSSSSVYQSVABGKMIWGLTYEDPSVNL 240
 LTNILLAKGSGYTN AM Y+K+L N+N+I++SSSVYQSVABGKMIWGLTYEDP +NL
 Sbjct: 181 LTNILLAKGSGYTNADAWAYMKLLVNMNSIRATSSSVYQSVABGKMIWGLTYEDPCINL 240

Query: 241 QKSGANVSIVPYTETCTVFPSSVAIKHAPSMKEAKLFINFMLSVDQNAFGQSTSNRPI 300
 QKSGANVSIVYP ECTVFPSSVAIK+AP+M EAKLFINFMLS DVQNAFGQSTSNRPI
 15 Sbjct: 241 QKSGANVSIVPYKSTVFPSSVAIKHAPNMTEAKLFINFMLSVDQNAFGQSTSNRPI 300

Query: 301 RKDAQTSNGCMALGDIALTKEDYRYVTKHGKQILKTYNRIRRNAD 345
 R+DAQTS+ MKAL+ IATLKEDY YVTKGK +I+ TYN++R+ +
 20 Sbjct: 301 RQDAQTSNDMALETIATLKEDYAYVTKHGKKKIVATTNQLRQRL 345

SEQ ID 1616 (GBS263) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 4; MW 63kDa).

The GBS263-GST fusion product was purified (Figure 205, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 301), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 506

A DNA sequence (GBSx0544) was identified in *S. galactiae* <SEQ ID 1619> which encodes the amino acid sequence <SEQ ID 1620>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4733 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF31452 GB:AF221126 putative response regulator [Streptococcus pneumoniae]
 Identities = 85/252 (33%), Positives = 147/252 (57%), Gaps = 17/252 (6%)

Query: 2 YRLILVEDEHLIRKMLRYAIDYQSLNIAVGEARDKGEKAQLIQECPDIVLSQIMPM 61
 Y +LIVEDE+L+R L ++ + + +G+A+G+ +LIQ+ PDI+L+IMPM
 45 Sbjct: 3 YTLILVEDEYLVKQGLTKLVNVAAYDMELIQAEINRQAWELIQVDPDIILTDIMPHL 62

Query: 62 TAFDMFEATKQGSYAK---IILSGYADFPNQAISHYGVLEPLTLEKQALIDCLKTIM 118
 + + +Y+ + +L+GY DF A SA+ GV +L KP +Q + + L I
 50 Sbjct: 63 NGIQLASLVR-ETYPQVHLVPLTGYDDFTYALSAKLVGDDYLLKPPSRQIEMLGKIK 121

Query: 119 ARIE-ESHEKILQETIELYLPQANDQVFEVTKDMANIESHPHGKIVISQLAHLGYS 177
 +++ R KE+ LQ+ L + + + I+ LA + + LA DLG+S
 Sbjct: 122 QKLDKEEKEQLQD-----LLNRFEGNNAQKIQSHLA-----DSQPSKSLASDLGFS 170

55 Query: 178 ESYLYTIVTKHILHITLSDYINQYRINQALQMFPRPDLAVYQIASAGVYDYRYFPRVFK 237
 +YL ++ KK L + DY+ + R+ QA +L+ DL +Y+IAE WG D YF + FK

-601-

Sbjct: 171 PTVLSSLIKKEGLGEPQDYLVRKVKQA-KILLITTDLKIYIEAKVGFDMMYFQRFK 229

Query: 238 KYLGQTVKAFKE 249

+ G T + FK+

Sbjct: 230 QTAGVTFRQFKK 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1621> which encodes the amino acid sequence <SEQ ID 1622>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 193/257 (75%), Positives = 226/257 (87%)

Query: 1 MYRLIVEDSHLRKWLRYTAIDYQSLNVLVGEARDGKEGAQLQESQPDIVLSDINMFI 60

MY+L+I+EDER+IRKWLRYTAIDY++L+ILV+GEARDGKEGA LI+E QPDIVL+DINMFI

Sbjct: 1 MYRLVIEDSHLRKWLRYTAIDYKALDILVIGENADGKEGAVLIKESQPDIVLTDINMFI 60

Query: 61 NTAFDMEFATKQGSYAKLLISGYADFFNAGSAHYGVLEFLTKFLEKQALIDCLKTIMAR 120

NTAFDMFE TK Q+YAKLLISGYADFFNAGSAHYGVLEFLTKP+EK AL +CL+TI+A+

Sbjct: 61 NTAFDMEFVTKDQYAKLLISGYADFFNARSAYHYGVLEFLTKFIEKALWECCLQTI IAK 120

Query: 121 IEHKEKHLQEHTELYLELPQANDQVPEVVKMLAWITHSHFHGKIVISQALNDLGYSESY 180

IE+ K + + +Y+ELFQ DQ+PEV+KDL N+H+HP KI S+LHDLGYSESY

Sbjct: 121 IEKQKSGMQKTACVYIFLPQMTDQIPEVVKDILEWVAHFQDKISTERLHDLGYSESY 180

Query: 181 LYTIVTKKHLNITLSDYINQYRINQAIQLM+REEDLMVYIAEAVGIYDYRYFDRVFKKYL 240

+Y KGH + LSDYINQYRINQAIQLM +BEDLMVY+IA+AVGIYDYRYFDRVFKKYL

Sbjct: 181 IYQNIKKHLQMLPSDYINQYRINQAIQLMQEEDLMVYIEAAGVYDYRYFDRVFKKYL 240

Query: 241 GQTVKAFKEEHIFKMD 257

GQTVKAFKEEH K D

Sbjct: 241 GQTVKAFKEEHFMKTD 257

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 507

A DNA sequence (GBSx0545) was identified in *S.agalactiae* <SEQ ID 1623> which encodes the amino acid sequence <SEQ ID 1624>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2964(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 508

A DNA sequence (GBSx0546) was identified in *S. agalactiae* <SEQ ID 1625> which encodes the amino acid sequence <SEQ ID 1626>. This protein is predicted to be two-component sensor histidine kinase.

Analysis of this protein sequence reveals the following:

```

5      Possible site: 45
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-13.80    Transmembrane    266 - 282 ( 257 - 285)
      INTEGRAL    Likelihood =-12.90    Transmembrane    29 - 45 ( 24 - 51)

10     ----- Final Results -----
           bacterial membrane --- Certainty=0.6519 (Affirmative) < succ>
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15     A related GBS nucleic acid sequence <SEQ ID 10197> which encodes amino acid sequence <SEQ ID
      10198> was also identified.
  
```

The protein has homology with the following sequences in the GENPEPT database:

```

      >GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
      [Bacillus halodurans]
20     Identities = 84/258 (32%), Positives = 138/258 (52%), Gaps = 23/258 (8%)

      Query: 298 SSAINGQVLDMDAISRQEKSSIELDSDQEPQYLSVQINQMVSRLKDLHEKTLDLSTQKLL 357
           S INQ+ S K+ I +D +DE LSVQ NGMV+ L+ L + + QK L
      Sbjct: 327 SERINQVA-----SGDLTKFKIVDGKDEIGQLSVQFNQMVANLRSLIHQVHETNRQRL 380

25     Query: 358 FEK-----RMLEAQFNPHFLNLTLETILITSHVDSQL-TERIVIGQTKLLAYSLSGST 409
           EK +ML +Q NHPL+NTLE+I + SH + ++V QD KL+R SL +
      Sbjct: 381 LEKSQNEIKMLASQINPHFLPYLTESIRKSHMKGETEIAKVVKQLGLMKKSLVETG 440

30     Query: 410 EAAVLKDDLAITESYLLINQVRF-BELTYTISVSPLEHMRVPLKFLPLLEIAIKYGLK 468
           L+++L ++ YL I R+ + L Y + + P+ E + + L + PL+ENA+ +GL+
      Sbjct: 441 HHPLRNELDWRCYLEIQTFYGRDLRYELYIDPQSEMVEILPLTIQPLVENAVINGLE 500

35     Query: 469 ERHD-VAINIDIWQSDGIWFTVSNNGSGISLARQQAIRIMLRSTH---SHHGLINSYR 523
           D + I + + + V+++G G+ + +AI+ ML + GL+N ++
      Sbjct: 501 RTEDGGTVITISTVNGDLITVINDDGCGMDEEKLEAIQNLHHPQEVGNKIGLLNVHK 560

      Query: 524 RLQVGF---STVLLEPTK 538
           RLQ + S +++E K
40     Sbjct: 561 RLQLTYGKTSGLITESAK 578
  
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1627> which encodes the amino acid sequence <SEQ ID 1628>. Analysis of this protein sequence reveals the following:

```

      Possible site: 43
45     >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-10.88    Transmembrane    27 - 43 ( 22 - 49)
      INTEGRAL    Likelihood = -9.08    Transmembrane    263 - 279 ( 258 - 282)

50     ----- Final Results -----
           bacterial membrane --- Certainty=0.5352 (Affirmative) < succ>
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
  
```

55 The protein has homology with the following sequences in the databases:

```

      >GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
      [Bacillus halodurans]
      Identities = 85/270 (31%), Positives = 139/270 (51%), Gaps = 20/270 (7%)
  
```


-603-

Query: 276 IFVILQRKSSGLANRIAAKNSRAINQMVDMISAISRQEKRRIDLESQDEFPQYLSQDQINQM 335
 + V+L S W L ++ + S INQ+ S K +I ++ +DE LS Q NQM
 Sbjct: 307 VAVLLIHVHFSMLSKRLSHLSERINQVA-----SGDLATKIVVDGKIQGLSVPQINQM 360

5 Query: 336 VERLQQLHDKTLLDLETQKLLFEK-----RMLAQPNPHFLNYTLETILITSHYDSAL- 387
 V L+ L + + QK L EK +ML +Q NPHFLNYTLE+I + SH
 Sbjct: 361 VAMLRSLTHQVHETNRQRRLLEKSQNETKLMLASQINPHFLNYTLESIRMKSHMKGETE 420

10 Query: 388 TEKIVQLTKLLRYSLTDSSEKPVLLKDDLESVIESYLVINQVRF-RRLQYSINLSFDLDSL 446
 K+V QL KL+R SL + + L+++L ++ YL I R+ + L Y + + P + +
 Sbjct: 421 IAKVQVQLGKLNKRSLEVTGHHIPLANELDMWRVCYLEIQFPFYGRLLHYELYIDPQSEM 480

15 Query: 447 EVPKFLFLPLIENAIKYLKERND-VKIMACYQDDHIFSVRONGSGIDAHQKVIRE 505
 E+ L + PL+ENA+ +GL+ D + I+ + + + V D+G G+D + I+
 Sbjct: 431 EILPLIIQPLVENAVIHGLERTEDGQTVHISTIVNGDLTVIVNMDGCGHDEKLEAIQN 540

20 Query: 506 QL---EAGESHGLINSYRRLKYHFEVS 531
 L E + GL+N ++RL+ + + S
 Sbjct: 541 MLRHQPEVDGNKIGLLNVHKLQLYTKIS 570

An alignment of the GAS and GBS proteins is shown below:

Identities = 369/549 (67%), Positives = 449/549 (81%)

25 Query: 3 MRGYRGEERFKRLQDDISKHFSRQSLIISLLLIALLFVLSLAPQIGLYKDVNSVSY 62
 MRG ++EE FKK+LQDDIS+HFS QSL+LSLLLI LF++FSLAPQ+GLY+D+N+ + Y
 Sbjct: 1 MRGGEYEEHFKKQLQDDISRRFSYSIMLSLLLIIGLFIIFSLAPQGLYLRDINATRY 60

30 Query: 63 KQLIKKHETLLDDLGKNSLKPFVSGHLSADLSKQYYHLRNLQSQTLLVFSPEQLLF 122
 +LI K + LDDLGKNSL PF++ +L +ADLSK Y+HLR+ Q+ ELL+FSF+Q+LLF
 Sbjct: 61 HRLISKQALLDDLGKNSLPLFNKNLSADLSKHYFHLRHSSQTSPELLLFSFQDLIF 120

35 Query: 123 ASNHLGNFFSKSIYISEVLDIAKINRLLKLIIVDSBGGHYLALIKPIIVNKVSGYAF 162
 ASN HGN FSKS+YI EVL + L K +DSE GHYL +I P+I ++ GYAF
 Sbjct: 121 ASNPHLGNVFSKSVYIQEVLRAETHSPKTLFKDAMDSBGGHYLMIIMPIDQMLGYAF 180

40 Query: 183 LMGKDFLLPTKAINSDLIADQLANSFTFNRDFISSLDKVDQSFLTRYFSFHDRAF 242
 +M+GKDFL PTK + S+L+IAD+L+N+FTF+NR+FI+SSLDK++S+L YF P D+RAF
 Sbjct: 181 VMGKDFLHPTKTLTSELVIADKLONTFTFSNREFIASSLDKINSQYLHHYFFQDNRAF 240

45 Query: 243 VVRKVALQDNILLMYRPLIPVTLVPLFELVSVIIPVILQESRVLADRIAVNBSAIN 302
 + RKVALQ + LMYRPLIP+ V+LPL+SS +IPVIL+ES LA+RIA KIS AIN
 Sbjct: 241 ITRKVALQGLWLMYRPLIPMVSVMLFELISSAVIPVILQRKSSGLANRIAAKNSRAIN 300

50 Query: 303 QNVLDMDAISRQKSEIHLSDQDEFOYLSQVQINQWRSRLDKHEKTLLETQKLLFEK 362
 QNV DM AISRQK I+L+SDQFOYLS QINQW RL+ LH+KTLLETQKLLFEK 360
 Sbjct: 301 QNVDMISAISRQKRRIDLESQDEFOYLEDQINQWRLQQLHDKTLLLETQKLLFEK 360

55 Query: 363 LEAQPNPHFLNYTLETILITSHYDSQLTRIVIQTKLLRYELSGSTEAVALKDDLAIE 422
 LEAQPNPHFLNYTLETILITSHYDS LTR+IVIQTKLLRYSLA S++ +LDDL+IE
 Sbjct: 361 LEAQPNPHFLNYTLETILITSHYDSALTRKIVIQTKLLRYSLTDSKPVLLKDDLESVIE 420

60 Query: 423 SYLLINQVRPEELITYTSVSPELHMVVKLFLEPLIENAIKYLKERHDVAINDIDWD 482
 SYL+INQVRPEL Y+I++SP+L+ + VKLFLEPLIENAIKYLKERHDV INI +
 Sbjct: 421 SYLVINQVRPELQYSINLSFDLSLEVPKFLLEPLIENAIKYLKERHDVKINACYQ 480

65 Query: 483 SDGIWPTVSNNGSGISLARQQAIRIMRSTSHHGLINSYRRLQYQFVTLLEFPKTD 542
 D I P+V +NNGSGI Q+ IR L+ + SHHGLINSYRELAY PS V L F + D
 Sbjct: 481 DDHIFSVRNGSGIDAHQKVIRQLAEGESHGLINSYRRLKYHFEVS LVPQDQ 540

Query: 543 FRVSYIVKE 551
 F VSY VKE
 Sbjct: 541 FNVSYIVKE 549

A related GBS gene <SEQ ID 8587> and protein <SEQ ID 8588> were also identified. Analysis of this protein sequence reveals the following:

```

Lipid: Possible site: -1      Crend: 10
McG: Discrim Score:         6.23
GVH: Signal Score (-7.5): -0.0500002
    Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value: -13.80 threshold: 0.0
  INTEGRAL    Likelihood =-13.80  Transmembrane  259 - 275 ( 250 - 278)
  PERIPHERAL  Likelihood =  2.70           404
modified ALOM score:  3.26

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.6519(Affirmative) < success
      bacterial outside --- Certainty=0.0000(Not Clear) < success
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the databases:

33.2/53.9% over 181aa
Streptococcus pneumoniae
GP|5830535| histidine kinase Insert characterized

ORF00032|1309 - 1848 of 2253
GP|5830535|emb|CB54576.1||A|0706396|1 - 182 of 231| histidine kinase [Streptococcus pneumoniae]
%Match = 5.9
%Identity = 33.2 %Similarity = 53.8
Matches = 61 Mismatches = 78 Conservative Sub.s = 38

[illegible][illegible]

1758 1788 1818 1848 1878 1908 1938 1968
IDTWQSDSGIWFTVNNNGSISLARQAQTMLRSTSHHGLINSYKRIQPQSTVLLEFTRTDADFVSIVTYSKE+VMYR
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
IKALQDGFVEILVDNRGMSAEKLINREKLSQRYPHQASYSYDQRSIGIVNHWFRVLYPGDRYAITYTSAPAGV
160 170 180 190 200 210

SEQ ID 8588 (GBS47) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 2; MW 84kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 4; MW 59.3kDa).

GBS47-His was purified as shown in Figure 221, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 509

A DNA sequence (GBSx0547) was identified in *S. agalactiae* <SEQ ID 1629> which encodes the amino acid sequence <SEQ ID 1630>. This protein is predicted to be phosphotransferase enzyme II, D component. Analysis of this protein sequence reveals the following:

```

5   Possible site: 32
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL   Likelihood = -10.46   Transmembrane 250 - 274 ( 252 - 274)
      INTEGRAL   Likelihood = -9.13    Transmembrane 232 - 248 ( 227 - 251)
      INTEGRAL   Likelihood = -5.31    Transmembrane 142 - 158 ( 140 - 161)
10   INTEGRAL   Likelihood = -2.50    Transmembrane 119 - 135 ( 118 - 139)

    ----- Final Results -----
          bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAC74689 GB:AB000276 PTS enzyme IID, mannose-specific
      [Escherichia coli K12]
20   Identities = 94/280 (33%), Positives = 156/280 (55%), Gaps = 13/280 (4%)

Query: 3   SQDNLTKEDEKMLRSVPWRSWTNASRTGATQYHAGVVIYTLFPVINRPFYKTDK-KAE 61
      ++ LT+ D +R VF RS S + A+G +++P I R Y + + + +A
25   Sb|ct: 12 TEKKLTQSD---IRGVFLRSLNLFQGS-WNFRMQLGFCTSMVPAIRRLTFENNARKQA 67

Query: 62   LVRETTWFNATHMHNIFIMGLVASMEEKNSDEPDFDASAITAVKASLMPGISGVDSFFW 121
      + EH +FN + I+G+ ++E++ + + D AI +K LMSR+ +GVGD FW
30   Sb|ct: 68 IRRHLEFFNTQPFVAAPILGVTLALEBQRANGAEIDDGAINGIKVGLMGPLAGVGFIFW 127

Query: 122  GILKVIAGGISLASTGSAMGAVFLLLYNIPAFILHYYSLNGSISVGAGFIKKLYESG 181
      G +R + A +G +A +GS +G ++F +L+N+ XY + GYS G +K + G
35   Sb|ct: 128 GTVPFVPAALGRGIAHSGSLGSLLEFIFLNLVRLVRIATRYGAYGSGKIDVKNM-GGG 186

Query: 182  GIKIVKTSSMLGLMVGSM----TASHVFKTILTVAAKGAKEASIQSYLDQLFVGTV 237
      ++ +7+ +S+LGL +G++ T N+ G + ++Q+ LDGL G+V
40   Sb|ct: 187 FLQKLTGRASILGLFVKGALVNRKWNHPIVVSRTDGTGKEHVTTVQTITDQLMPLV 246

Query: 238  PLAVTILAFMLLRKKVNNINMFGIMVLGI---VLGLGI 274
      FLL+T WLLRKKVN WI+ G V+GI GLG+
45   Sb|ct: 247 PLLLTFACMWLLRKKVNPLMIIVGPFVIGIAGYACGLLGL 286

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1631> which encodes the amino acid sequence <SEQ ID 1632>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 32
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL   Likelihood = -9.98    Transmembrane 255 - 271 ( 251 - 274)
      INTEGRAL   Likelihood = -7.01    Transmembrane 232 - 248 ( 228 - 250)
      INTEGRAL   Likelihood = -5.68    Transmembrane 142 - 158 ( 140 - 161)
50   INTEGRAL   Likelihood = -2.50    Transmembrane 119 - 135 ( 118 - 139)

    ----- Final Results -----
          bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAC74689 GB:AB000276 PTS enzyme IID, mannose-specific
      [Escherichia coli]
60   Identities = 94/280 (33%), Positives = 157/281 (55%), Gaps = 13/281 (4%)

Query: 2   TSQDNLTKEDEKMLRSVPWRSWTNASRTGATQYHAGVVIYTLFPVINRPFYKTDK-KAE 60

```

-606-

T++ Lf+ D +R VF RS S + A+G ++++P I R Y + + + +
 Sbjct: 11 TTEKLTQSD---TRGVFLASNLFGQS-WNPERMAGLFCPSWVAIRRLYPENNEARQ 66
 5 Query: 61 ALVRHTTWFNATMHIINFMELVASMEEKKNSDEDFDASAITAVKASLMSPISGVGDSFF 120
 A+ RH -FN + I+G+ ++E++ + + D AI 4K LMEP++GVGD F
 Sbjct: 67 AIRRHLEFNTQPPVFAAPILGVTLALSGQRANGAEIDEGALNGIKVGLMGFLAGVGDP 126
 10 Query: 121 WGLLEVIAAGIGISLASAGSANGAVFVLLYNIAPFIHYSLYGGYSVGAQFIKKLYES 180
 WG +R + A +G +A +GS +G +F +L+H+ Y+ + GYS G + K +
 Sbjct: 127 WGVFVFPFALGAGIAMSGLLGLPLFFLLFNLVRLATRIYGVAYGYSKIDIVKDM-GG 185
 15 Query: 181 GGIKIVTKTSSMLGLMMVGS- ---TASNVKFKTILVAAKGAKEAASIQDYLQPLFI 236
 G + +T+ +S+LGL ++G+ T H+ G + +Q LDQL G+
 Sbjct: 186 GFLQKLTGASILGLFVKGALVKNKTHVNIPLVSRITDQTKSHVPTVQTILDQMLPGL 245
 20 Query: 237 VFLMVTIAAFVLLRKKVNIWIMPGIMFLGI---ILGLGI 274
 VFL+T+ A WLLRKKVN +W+ G +GI GLG+
 Sbjct: 246 VFLLLTFACWLLRKKVNIWIMPGIFVFGVIGIAGYACGLLGL 286

An alignment of the GAS and GBS proteins is shown below:

Identities = 263/275 (95%), Positives = 269/275 (97%)

Query: 1 MTSQNLTYKEDRIKRLSVFWRGWINNASRTGATQYHAGVVIYTLFVFNRFKTDKDAE 60
 M SQNLTYKEDRIKRLSVFWRGWINNASRTGATQYHAGVVIYTLFVFNRFKTDKDAE
 25 Sbjct: 1 MTSQNLTYKEDRIKRLSVFWRGWINNASRTGATQYHAGVVIYTLFVFNRFKTDKDAE 60
 Query: 61 ALVRHTTWFNATMHIINFMELVASMEEKKNSDEDFDASAITAVKASLMSPISGVGDSFF 120
 ALVRHTTWFNATMHIINFMELVASMEEKKNSDEDFDASAITAVKASLMSPISGVGDSFF
 30 Sbjct: 61 ALVRHTTWFNATMHIINFMELVASMEEKKNSDEDFDASAITAVKASLMSPISGVGDSFF 120
 Query: 121 WGLLEVIAAGIGISLASAGSANGAVFVLLYNIAPFIHYSLYGGYSVGAQFIKKLYES 180
 WGLLEVIAAGIGISLAS GSKMAGVFLLLYNIAPFIHYSLYGGYSVGAQFIKKLYES
 Sbjct: 121 WGLLEVIAAGIGISLASAGSANGAVFVLLYNIAPFIHYSLYGGYSVGAQFIKKLYES 180
 35 Query: 181 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILVAAKGAKEAASIQDYLQPLFI 240
 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILVAAKGAKEAASIQDYLQPLFI
 Sbjct: 181 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILVAAKGAKEAASIQDYLQPLFI 240
 40 Query: 241 VTIAAFVLLRKKVNIWIMPGIMFLGIILGLLGIC 275
 VT+ AFVLLRKKVNI WIMPGIM LQI+LGLLGIC
 Sbjct: 241 VTIAAFVLLRKKVNIWIMPGIMFLGIILGLLGIC 275

There is also homology to SEQ ID 5236.

A further related DNA sequence was identified in *S. pyogenes* <SEQ ID 9077> which encodes the amino acid sequence <SEQ ID 9078>. An alignment of the GAS and GBS sequences follows:

Score = 178 bits (448), Expect = 3e-47
 Identities = 93/136 (61%), Positives = 108/136 (79%)
 50 Query: 2 IMESITTYHNPCGTSRNVLAIRHAGIEPTTIEVLOTTPNRITIELQSGISARELL 61
 +ME+I IYHNPCGTSRNVLA+IRH GIEP II YL+TP+R L+ELL N +SARELL
 Sbjct: 1 MMEKIRIYHNPCGTSRNVLAIRHAGIEPTTIEVLOTTPNRITIELQSGISARELL 60
 55 Query: 62 RTDVPEPEAYGLANQAVAEKDIINAMLDPIILNRPVIVTKGVKLCRPGFTLLDLPVP 121
 RT+VE +E + L + +V ++++I+AM+ DPILNRPVIVT KG KLCRP E +L ILVP
 Sbjct: 61 RTDVPEAYEKNLESSSVTDEEMIDMIQDPILNRPVIVTKGVKLCRPGFTLLDLPVP 120
 Query: 122 LPSFYIKEDGESVNDI 137
 + ++KEDG+ + +
 Sbjct: 121 MEKDFVKEDSQIIQSL 136
 60

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 510

A DNA sequence (GBSx0548) was identified in *S.agalactiae* <SEQ ID 1633> which encodes the amino acid sequence <SEQ ID 1634>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

```

5   Possible site: 56
   >>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70    Transmembrane 144 - 160 ( 140 - 165)
   INTEGRAL    Likelihood = -8.07    Transmembrane 220 - 236 ( 215 - 239)
10  INTEGRAL    Likelihood = -7.27    Transmembrane 95 - 111 ( 91 - 116)
   INTEGRAL    Likelihood = -3.77    Transmembrane 2 - 18 ( 1 - 18)
   INTEGRAL    Likelihood = -1.44    Transmembrane 180 - 196 ( 179 - 196)
   INTEGRAL    Likelihood = -1.33    Transmembrane 32 - 48 ( 30 - 49)
   INTEGRAL    Likelihood = -0.53    Transmembrane 198 - 214 ( 198 - 214)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20  The protein has homology with the following sequences in the GENPEPT database:
   >GP:NAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
   [Vibrio furnissii]
   Identities = 70/251 (27%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

25  Query: 2 IMPATMAALAVLICFOGNYLTGQSMERPLVVLGVLGMLGDIKVLGMASLEALFLQN 61
   + A M L + G + G + RP+V+G + G+LGD+ GIL+G +LE ++G
   Sbjct: 5 LFGALMLGLLAFLA-GLDLFNLGTHFRFRPVLGPIVGLILGDLHTSLVGOTLELIWMLG 63

30  Query: 62 VNIIGGVIAEPVVTATAMATTTIISNIDQKAMTLAVFIQMLAAFPVPLKVNFPNIPAP 121
   + G + T + TTF I +N++ R+ +RVE + + L + + +
   Sbjct: 64 APLAGAQPPNVIIGTIGTTFITATTNVEPNVAGVAVFFAVAVQMGITLLPSAMSAVMK 123

35  Query: 122 MVDKAAANHQGKLVMHYGTWII--YYLIASISFIIIGVSGFVNSFVHHIPQNLQNG 179
   + A R+ +G + ++ ++ +Y + A P+ I +G+ + V +P+ L++G
   Sbjct: 124 CDEYAKNADTRGIERVNYPALAVLGSFYFLCA---PLPTYLGDADNAGMVAALPKALIDG 180

40  Query: 180 LSAAGGLLPAVGFMKMLMLWTNKLAVFYLLGFVLTAYLKLPAAVAALGAVICVISSQR 239
   L AGG++PA+GFA+LMK++ N +++LGFV A+L+LE +R+ + +I R
   Sbjct: 181 LGVAGIMPALGFAVLMKIMMKNAYIYFYLLGFVAAALQLQLILARCAATAMATIDFPR 240

45  Query: 240 DIELDATITRGA 250
   + + A
   Sbjct: 241 KSEPTFVNAGA 251

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1635> which encodes the amino acid sequence <SEQ ID 1636>. Analysis of this protein sequence reveals the following:

```

   Possible site: 56
   >>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70    Transmembrane 144 - 160 ( 140 - 165)
50  INTEGRAL    Likelihood = -8.07    Transmembrane 220 - 236 ( 215 - 239)
   INTEGRAL    Likelihood = -7.27    Transmembrane 95 - 111 ( 91 - 116)
   INTEGRAL    Likelihood = -4.62    Transmembrane 2 - 18 ( 1 - 19)
   INTEGRAL    Likelihood = -1.44    Transmembrane 180 - 196 ( 179 - 196)
   INTEGRAL    Likelihood = -0.96    Transmembrane 32 - 48 ( 31 - 49)
55  INTEGRAL    Likelihood = -0.53    Transmembrane 198 - 214 ( 198 - 214)

   ----- Final Results -----
        bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-608-

The protein has homology with the following sequences in the databases:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIFMan
(Vibrio furnissii)
Identities = 72/251 (28%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

5 Query: 2 LVEATMAALAVLICFGGNYLTGQSMMERPLVVGVLVTGLLGDIMKVGILMGASLEALPLGN 61
  L A M L + G + G + R+V+G + GL+LGD+ GIL+G +LE ++G
Sbjct: 5 LPAALMLGLLAFLA-GLDLPNGLTHPRPVLGFLVGLILGDLATGILVGGTLEILVMGL 63

10 Query: 62 VNIGGVIAAEPTATAMATTFTIISHIDQKAMTLAVPIGMLAFAFVPMFLKNVPMIIFAP 121
  + G + T + TTF I +++ A+AVP + + L+ + +
Sbjct: 64 APLGAQPPVVIOTIVGTTFAITINVEPNVAVGVAVPFAVAVQMGITLLPSMSASVMSK 123

15 Query: 122 MVDKAAAANHQGLVLMHLYGTWII--YYLIIASISFIGILVSGSPVNAFVEHIPQNLNGL 179
  + A A+ +G + + + ++ +Y + A F+ I +G+ A V +P+ L+ +G
Sbjct: 124 CDEYAKGADTRGIERVNYFALAVLGSFYFLCA---FLPTILGADHAGMVAALPKALIDG 180

Query: 180 LSAAGLLPAVGFAFAMLNKLLWTKLAVFYLLGFVLTAYLKLPAAVAAALGAVICVISSQR 239
  L AGG+ +A+FA+LANK++ N +++LGFV A+L+LP+A+ + I R
20 Sbjct: 181 LGVAGGIMEALGFVAVLNKIMMKVAYIPYFILGFVAAAMLQLPILARCATAVAIIDFWR 240

Query: 240 LELDAITRGA 250
  E + A
25 Sbjct: 241 KSEPTFWASA 251
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 261/269 (97%), Positives = 268/269 (99%)

30 Query: 1 MIMPATMAALAVLICFGGNYLTGQSMMERPLVVGVLVTGLLGDIMKVGILMGASLEALPLG 60
  M++PATMAALAVLICFGGNYLTGQSMMERPLVVGVLVTGLLGDIMKVGILMGASLEALPLG
Sbjct: 1 MLVPATMAALAVLICFGGNYLTGQSMMERPLVVGVLVTGLLGDIMKVGILMGASLEALPLG 60

Query: 61 NVNIGGVIAAEPTATAMATTFTIISHIDQKAMTLAVPIGMLAFAFVPMFLKNVPMIIFAP 120
  NVNIGGVIAAEPTATAMATTFTIIS+IDQKAMTLAVPIGMLAFAFVPMFLKNVPMIIFAP
35 Sbjct: 61 NVNIGGVIAAEPTATAMATTFTIISHIDQKAMTLAVPIGMLAFAFVPMFLKNVPMIIFAP 120

Query: 121 MVDKAAAANHQGLVLMHLYGTWIIYYLIIASISFIGILVSGSPVNAFVEHIPQNLNGL 180
  MVDKAAAANHQGLVLMHLYGTWIIYYLIIASISFIGILVSGSPVNAFV HIPQNLNGL
40 Sbjct: 121 MVDKAAAANHQGLVLMHLYGTWIIYYLIIASISFIGILVSGSPVNAFVEHIPQNLNGL 180

Query: 181 SAAGGLLPAVGFAFAMLNKLLWTKLAVFYLLGFVLTAYLKLPAAVAAALGAVICVISSQRD 240
  SAAGGLLPAVGFAFAMLNKLLWTKLAVFYLLGFVLTAYLKLPAAVAAALGAVICVISSQRD
45 Sbjct: 181 SAAGGLLPAVGFAFAMLNKLLWTKLAVFYLLGFVLTAYLKLPAAVAAALGAVICVISSQRD 240

Query: 241 LELDAITRGAISKOTTFDKSSEERDFFA 269
  +ELDAITRGAISKOTTFDKSSEERDFFA
Sbjct: 241 LELDAITRGAISKOTTFDKSSEERDFFA 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 511

A DNA sequence (GBSx0549) was identified in *S. agalactiae* <SEQ ID 1637> which encodes the amino acid sequence <SEQ ID 1638>. This protein is predicted to be pts system, sorbose-specific iib component. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1874 (Affirmative) < succ>
60 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
```

-609-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:CAA6858 GB:X66059 EIII-B Sor PTS [Klebsiella pneumoniae]
 Identities = 49/158 (31%), Positives = 94/158 (59%), Gaps = 8/158 (5%)

Query: 2 ITQIRVDDRLLHSGVAVVWTKENAPLLVAVNDEAAKNEITQMTLQAVVNGMKLLIRS 61
 IT R+DDRLLHSGV VV+K NA +++ ND+ +E+ + L+ A P GMR+ + S+
 10 Sbjct: 3 ITTARIDRLLHSGVTTVNSKVANAQRITTCNDVFNDEVRITLGRQAAPPGRKVVVNSL 62

Query: 62 RESIALFKDPRAITDKRIFVIVNSVKDACFIANKITDLEAVNVANVGRFDSDPATKVKLT 121
 E+++A++ +P+ D+ +P+ + D T+ + + +H+ + + K +LT
 15 Sbjct: 63 EKAAVAVYHNPQYQUBTVFYLFYFNPHDVLITMVGRQGVQIATINIGGM-----AWRPGKQLT 117

Query: 122 SSSLINTELEAAKELASL-PDLDFVFNQVLSPTKVN 158
 ++ Lt ++++A +EL L LD+ +V+ S+ VN+
 Sbjct: 118 KAVSLDFQDIQAFKELDKLSVKLDL--RVVASDFSVNI 153

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1639> which encodes the amino acid
 20 sequence <SEQ ID 1640>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.1874 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 145/162 (89%), Positives = 152/162 (93%)

Query: 1 MITQIRVDDRLLHSGVAVVWTKENAPLLVAVNDEAAKNEITQMTLQAVVNGMKLLIRS 60
 MITQIRVDDRLLHSGVAVVWTKENAPLLVAVNDEAAKNEITQMTLQAVVNGMKLLIRS
 35 Sbjct: 1 MITQIRVDDRLLHSGVAVVWTKENAPLLVAVNDEAAKNEITQMTLQAVVNGMKLLIRS 60

Query: 61 VESSIALFKDPRAITDKRIFVIVNSVKDACFIANKITDLEAVNVANVGRFDSDPATKVKLT 120
 VE+GI LF DPRA DKRIFVIVNSVKDAC IAK + DLEAVNVANVGRFDSKDPH+KVK+
 Sbjct: 61 VEDSILKFNDFRAIDKRIFFVIVNSVKDCAIAKESVPDLEAVNVANVGRFDSKDPKASKVIV 120

40 Query: 121 TSSLLINTELEAAKELASLPDLDFVFNQVLSPTKVNLSQLV 162
 T SLLIN RE+ AKKL SLF+LDVFNQVLSPTKVN+LSQLV
 Sbjct: 121 TPSLLINPEEMAAKELVSLPELDFVFNQVLSPTKVNLSQLV 162

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 512

A DNA sequence (GBSx0550) was identified in *S.agalactiae* <SEQ ID 1641> which encodes the amino
 acid sequence <SEQ ID 1642>. Analysis of this protein sequence reveals the following:

possible site: 46
 50 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.22 Transmembrane 87 - 103 (87 - 104)

----- Final Results -----
 bacterial membrane --- Certainty=0.1489 (Affirmative) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1643> which encodes the amino acid sequence <SEQ ID 1644>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.44    Transmembrane    87 - 103 ( 87 - 104)

----- Final Results -----
10      bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```

15  Identities = 115/141 (81%), Positives = 125/141 (88%)

Query: 1  MRKKFLIGSHGKRLASGLQSSIDILTKGQCEIQTIDAYIDGSDYTKSIVEFIDEIAPDEQG 60
      MRKKFLIGSHGRLASGLQSSIDIL G GQ ++TIDAY+DSDYTT I +FI +A DEQG
20  Sbjct: 1  MRKKFLIGSHGRLASGLQSSIDILAGMGQALETIDAYVDDSDYTSQIDDFIAGVADEQG 60

Query: 61  LIPTDLLQGSVWQKMATVWNSGNDNIPLITNSHLATLLSLFLKPEEHLTKKEIIVTVIN 120
      LIPTDLLQGSVWQKMATVWNSGK+NIPLITNSHLATLLSL+FLKP E LTK+EIVTVIN
20  Sbjct: 61  LIPTDLLQGSVWQKMATVWNSGNDNIPLITNSHLATLLSLVFLKPGALTKDEIIVTVIN 120

Query: 121 ESQVQLVDLSPKAGSEDDFFD 141
      ESQVQLVDL + SEDDFD
25  Sbjct: 121 ESQVQLVDLVPETNSDDFFD 141

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 513

A DNA sequence (GBSx0551) was identified in *S.agalactiae* <SEQ ID 1645> which encodes the amino acid sequence <SEQ ID 1646>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
35  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2469 (Affirmative) < succ>
40      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 514

A DNA sequence (GBSx0552) was identified in *S.agalactiae* <SEQ ID 1647> which encodes the amino acid sequence <SEQ ID 1648>. This protein is predicted to be racemase. Analysis of this protein sequence reveals the following:

```

50  Possible site: 41
    >>> Seems to have no N-terminal signal sequence

```


-611-

```

5      INTEGRAL    Likelihood = -8.65      Transmembrane  319 - 335 ( 316 - 339)
      INTEGRAL    Likelihood = -6.10      Transmembrane  18 - 34 ( 17 - 37)
      INTEGRAL    Likelihood = -5.68      Transmembrane  230 - 246 ( 227 - 248)
      INTEGRAL    Likelihood = -3.98      Transmembrane  254 - 270 ( 254 - 271)
      INTEGRAL    Likelihood = -3.56      Transmembrane  110 - 126 ( 110 - 129)
      INTEGRAL    Likelihood = -3.19      Transmembrane  161 - 177 ( 156 - 177)
      INTEGRAL    Likelihood = -1.97      Transmembrane  132 - 148 ( 132 - 153)
      INTEGRAL    Likelihood = -1.33      Transmembrane  286 - 302 ( 286 - 302)
10     INTEGRAL    Likelihood = -0.59      Transmembrane  53 - 69 ( 52 - 69)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GF:AA71283 GB:AF253562 racemase [Enterococcus faecalis]
Identities = 78/262 (29%), Positives = 129/262 (48%), Gaps = 29/262 (11%)

20     Query: 13  KQRTSMISMLQYLFSLVILWISGRFPS-QDVIFHTKSLGRMAVFFLICITAFPLRG 71
      K + S I +++ ++L++ +H+ LPS + +F F + +NVP+P + + FEL
      Sbjct: 3    KNESYSGIDYFRFLAALLVAIETSLFPSSTGNFIFTRIVAPVAVFPFMTSGPFL-- 60

25     Query: 72  RIQQCLNHSYFRKLKK----YSMTIITLYLP----GYFFFSMLIAKIVLLGQFIVAF 123
      I + CN IKK Y + ++N+P GYF ++L LP I
      Sbjct: 61  -ISRYTCAEKLGAPIKKTITLYGVALLNIPINVYNGYFQDNL-----LEMIKKDI 112

30     Query: 124 LNLGNSHTLWYIPAVILGNVLIQGLLYGFRGTFTTVVVLNIGAV-ETYSVEIQSTKF 182
      ++ G + LNY+PA I+G I L+K V R P+ +IN IG ++Y ++S
      Sbjct: 113 VFDGTLVHLNVLPASTIGAAIWLKVKVYRJAFLASILNLIIGLPDSYGYGVKVSVC 172

35     Query: 183 YPLSTYMSYIFQT---TRNLEFTEPVVLLAGVLYDYFNTDLPTKSRGLK-YILPLLLA 238
      L Y IPQ TRN+P+ P++ + G + D + + + K ++ Y LP L+
      Sbjct: 173 --LNVFYINLI PQMDYTRNGSIFPAPFPVFGYISD--SPNRYRKNQVIRIYSLFCLMFG 228

      Query: 239 LENVLIYFN-CGLDNFPLIAP 259
      L +P+ Q D + LL P
      Sbjct: 229 KTLTLQHPDIQKIDSMYVLLLF 250

```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8589> and protein <SEQ ID 8590> were also identified. Analysis of this protein sequence reveals the following:

```

      Lipop: Possible site: -1  Crend: 7
      McG: Discrim Score: 0.23
45     GvH: Signal Score (-7.5): -5.77
      Possible site: 34
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 3 value: -5.68 threshold: 0.0
      INTEGRAL    Likelihood = -5.68      Transmembrane  41 - 57 ( 38 - 59)
      INTEGRAL    Likelihood = -3.98      Transmembrane  65 - 81 ( 65 - 82)
50     INTEGRAL    Likelihood = -1.33      Transmembrane  97 - 113 ( 97 - 113)
      PERIPHERAL    Likelihood = 5.78
      modified ALOM score: 1.64

55     *** Reasoning Step: 3

      ----- Final Results -----
      bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS gene <SEQ ID 8591> and protein <SEQ ID 8592> were also identified. Analysis of this protein sequence reveals the following:

Example 515

A DNA sequence (GBSx0553) was identified in *S.agalactiae* <SEQ ID 1649> which encodes the amino acid sequence <SEQ ID 1650>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3088 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 516

A DNA sequence (GBSx0554) was identified in *S.agalactiae* <SEQ ID 1651> which encodes the amino acid sequence <SEQ ID 1652>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1446 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 517

A DNA sequence (GBSx0555) was identified in *S.agalactiae* <SEQ ID 1653> which encodes the amino acid sequence <SEQ ID 1654>. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 10
MoG: Discrim Score: 8.28
GVH: Signal Score (-7.5): -2.11
Possible site: 20
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 6 value: -8.33 threshold: 0.0
40  INTEGRAL Likelihood = -8.33 Transmembrane 358 - 374 ( 354 - 376)
    INTEGRAL Likelihood = -8.23 Transmembrane 264 - 280 ( 257 - 290)
    INTEGRAL Likelihood = -6.37 Transmembrane 210 - 226 ( 206 - 232)
    INTEGRAL Likelihood = -5.95 Transmembrane 163 - 179 ( 160 - 180)
    INTEGRAL Likelihood = -5.10 Transmembrane 23 - 39 ( 21 - 40)
45  INTEGRAL Likelihood = -1.70 Transmembrane 297 - 313 ( 296 - 314)
    PERIPHERAL Likelihood = 1.75 322
    modified ALOM score: 2.17

*** Reasoning Step: 3

----- Final Results -----

```

-614-

```

bacterial membrane --- Certainty=0.4333 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 518

10 A DNA sequence (GBSx0556) was identified in *S.galactiae* <SEQ ID 1655> which encodes the amino acid sequence <SEQ ID 1656>. This protein is predicted to be ABC transporter (ATP-binding prot). Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.1510 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

20 A related GBS nucleic acid sequence <SEQ ID 10199> which encodes amino acid sequence <SEQ ID 10200> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAE88481 GB:AL353816 putative ABC transport system ATP-binding
protein [Streptomyces coelicolor A3(2)]
Identities = 104/284 (36%), Positives = 159/284 (55%), Gaps = 18/284 (6%)

Query: 6 TMLQLDNITKSYGKIVIANQISYQFTPLGLYGLLGANTGTGKTLNLMSEHTLADSGNIY 65
      + ++ YG+ L+ +S + TP+ QLG NG GKTL ++ AD G
Sbjct: 2 TPTVSQSLSHYGRTRALDVSRLTPGVTGLLGANGAKITLLRLVATVPADRGAPT 61

Query: 66 WNGQSQS-----EEFYRHIGFLPQHFYRYDQPTGIAPLNVIATLKGV-DKKAKQRIPEL 119
      G + +E R +G+LPQ ++ FT F+Y+A LK + D+++ +E+ R+
Sbjct: 62 VLGHDPGSSRGQRQEVRRLLGYLPQTGPHEPDTAFEFVDYVAILKELADRRHRHREVRV 121

Query: 120 LELVGLGDVGKKIKISSYSGMKQRLGIAQALINDPEILILDEPTVGLDPKERVKFRHLS 179
      LE V LG+V ++I SGM+OR+ +A AL+ DP L+LDEPTVGLDP++R++FR +++
Sbjct: 122 LEEVDLGEVGRRIKLSGCMQRQVALAALVGDGPFLVLEDEPTVGLDPQRMRELIA 181

Query: 180 QLSTNKIIILSTHIVSDVENAKREIVLKNQKPTIEKNTAQLLKTIGKVNREIT-TEPGL 238
      + ++LSTH DV + +IV+ G G A+L G+V T +PG
Sbjct: 182 GAGEGRIVLLSTHQTEDVAMLCRVIVMAGAVRFDGTFAILARAAGRVNSSTEDKPG- 240

Query: 239 SQIPNIAIVNEKVFSDSRVFRVSDICPSDSQILVPTLEDVFI 282
      A + + S FR V D P A+ PTLED Y+
Sbjct: 241 -----AKAGNRITGTS--FRNVGD--PPFGARPAETLEDGYL 274

```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 519

A DNA sequence (GBSx0557) was identified in *S. agalactiae* <SEQ ID 1657> which encodes the amino acid sequence <SEQ ID 1658>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```

5   Possible site: 38
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.3781 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:CA10170 GB:AJ278301 response regulator [Streptococcus pneumoniae]
   Identities = 136/242 (56%), Positives = 183/242 (75%)

   Query: 1 MNIFILEDQFVQQAHPFKIIRVQYNLHFKTVETPAKPVQLLESIEYIGLHNLFFLDI 60
   Sbjct: 1 MRIFVLEDDFSQQTRIETITLKLKEHHITLSSFEVFGKPDQLAEVHEKGAKHLLFFLDI 60

20  Query: 61 EIKNDEQMGLEVAQIRQVDPAQIVFVTTSSLMPLTFRYQVSALDYIDKGLSQEFSQ 120
   EI+N+E GLEVA++IR+ DPYA IVFVTTSS MFL+FRYQVSALDYIDK LS EEF
   Sbjct: 61 EIRNEEMKGLVARKIRQDPYALIVFVTTSSFMPLSFYQVSALDYIDKALSSEFES 120

25  Query: 121 RIEEVLLYVDGICNKLPLVENSFPFKSRYSQVQLFPNDLLIYETSSSRHVVLYTEKDRME 180
   RIE LLY + +K L Z+ FYPKS+++Q Q PF ++ Y+ETS R HRV+LYT+ DR+E
   Sbjct: 121 RIETALLYANSQDKSLAEDCFYFKSFAQYPPKSVTYLETSPRHVRVLYTKDRLE 180

30  Query: 181 FTATLGDILKQEPRLQCHRSFLNPLNIFKVDRIDRLVYPQNGTICLVSRNKVRDIVSI 240
   FTA+L ++ KQEPRL QCHRSFL+NF N+ +D+ ++L+F NG +CL+R KVR++
   Sbjct: 181 FTASLEEVFKQEPRLQCHRSFLINPANNVHLDKKELFFPNQGSCLIAIKYREVSEA 240

   Query: 241 VD 242
35  Sbjct: 241 IN 242

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1659> which encodes the amino acid sequence <SEQ ID 1660>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 44
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
45  bacterial cytoplasm --- Certainty=0.2098 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

   Identities = 106/235 (45%), Positives = 159/235 (67%)

50  Query: 1 MNIFILEDQFVQQAHPFKIIRVQYNLHFKTVETPAKPVQLLESIEYIGLHNLFFLDI 60
   Sbjct: 2 MNIFILEDQF+QQ E I+ I + + +E F+ P +L RSI E G H L+FLDI
   MNIFILEDQF+IQQTIESIVGILKTKHLPQNQLFVSTPQKLFSSTQERGDHLYFLDI 61

55  Query: 61 EIKNDEQMGLEVAQIRQVDPAQIVFVTTSSLMPLTFRYQVSALDYIDKGLSQEFSQ 120
   EI + GLE+A IRQ DP A IVFVTTSS P++F+Y+VSALD+IDK Q++F +
   Sbjct: 62 EIGETRCSELAALTRQKDPNAVIVFVTTSSFAPISTFKYKVSALDIDKAGGQKQKPE 121

   Query: 121 RIEEVLLYVDGICNKLPLVENSFPFKSRYSQVQLFPNDLLIYETSSSRHVVLYTEKDRME 180
   +IEE + Y + + + + F F++ ++++LP+ D+LY T++ H+V L+T+ +R+E
60  Sbjct: 122 QIEECIRYTYDMSRESKDWFLPETPQRLKLPQKLYLPATATTPHKVCWTQVSRLE 181

```

Query: 161 FTATLGDLAQEPRFLQCHRSLNPLNIFKVRIDRLVYFQNGTCLVSRNKVR 235
 F L +I P+LF CHRSLVN + ++D+ +L+YF+NG +C+VSR K++
 Sbjct: 162 FYGNLSEIQVAFKLFCHRSLVNLDKVVRIDKSKQLLYFENKDCMVSRLEWK 236

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 520

A DNA sequence (GBSx0558) was identified in *S. agalactiae* <SEQ ID 1661> which encodes the amino acid sequence <SEQ ID 1662>. Analysis of this protein sequence reveals the following:

10 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
 15 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1663> which encodes the amino acid sequence <SEQ ID 1664>. Analysis of this protein sequence reveals the following:

20 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0535(Affirmative) < succ>
 25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 177/269 (65%), Positives = 219/269 (80%)

Query: 6 MAKCLTLNTHSWMEVNAKKLFDLAERHIFREKYDIIICLOEVWQSISSPLAKSSPNWHPTE 65
 M K LTLNTHSWM+ N LKKL LAERH EKYDIIICLOEQ I S LA P Y +
 35 Sbjct: 1 MTKVTLNTHSWMQANTLKKLVALAEHTLAERKYDIIICLOEQINQLRESLATDLPRQALS 60

Query: 66 GTPALHQDNFALQVHYLMQLHYHWTWANNHIGSKYHEGVAILSLKPLKPEDILVSA 125
 GTP+H+D+FA L LHYL +G HY+W+HANNHIGY Y EGVAILS +P+ DILVSA
 35 Sbjct: 61 GTPSIHKDHFALLIHYLQKRGCHYTSWAYNHHIGYDYOEGVAILSKQPIHVSCLVSA 120

Query: 126 VDDETVDYHTRRALVAETTLNDKVVTVVSLHPSNPEKGFABEMKRLTTLLSEVETPLLNG 185
 +DDEYDYHTRR+L+A+TTL+ K V VV++H SNF+KGF EW++LE LL+ PLLNG
 40 Sbjct: 121 MDETDYHTRRSLIAKTTTLDGKEVAVVNVHLSNFDKGFLEWMEKLEKELLTIANCPILLNG 180

Query: 186 DENNPTNGQCVLEVLNLSPLAKLDSHQIANVHVDHTIMADIDGWEQNKALKVVDHFTSE 245
 DENNPT GY+ +++ SPL L+DSH+ A+HVSQDH+I+ADIDGW+GNK+ALKVDH+FTS+
 45 Sbjct: 161 DENNPTDQGCYQVWMSGPLDQSHKGAHIVFGDHSIVADIDGWQGNKALKVHVPFTSK 240

Query: 246 DLGISSSQCVFPEGQDAFVSDHYGLEITM 274
 D I SS++ FEGG+APVSDHYGLE+T+
 50 Sbjct: 241 DPIIRSSKITFEGGDAFVSDHYGLEVTL 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 521

A DNA sequence (GBSx0559) was identified in *S. agalactiae* <SEQ ID 1665> which encodes the amino acid sequence <SEQ ID 1666>. This protein is predicted to be PTS system, glucose-specific enzyme II, A component (ptsG). Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.07	Transmembrane	193 - 209 (189 - 217)
INTEGRAL	Likelihood = -7.86	Transmembrane	28 - 44 (24 - 48)
INTEGRAL	Likelihood = -6.48	Transmembrane	431 - 447 (421 - 449)
INTEGRAL	Likelihood = -2.92	Transmembrane	153 - 169 (153 - 170)
INTEGRAL	Likelihood = -2.81	Transmembrane	93 - 109 (93 - 111)
INTEGRAL	Likelihood = -2.39	Transmembrane	370 - 386 (370 - 388)
INTEGRAL	Likelihood = -2.28	Transmembrane	68 - 84 (68 - 84)

----- Final Results -----
 bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10201> which encodes amino acid sequence <SEQ ID 10202> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 294/409 (71%), Positives = 342/409 (82%), Gaps = 7/409 (1%)

Query: 293 DLINLKGS-NSSQVHLLTSVTPARFKVQMGIGSGILML+AMTRND DKK KTK M
 DL+LKG+ + SQVHLLTSVTPARFKVQMGIGSGILML+AMTRND DKK KTK M
 Sbjct: 3 DLHLKGRMSQVHLLTSVTPARFKVQMGIGSGILML+AMTRND DKK KTK M 62

Query: 352 FISAAAATFLTGVTPEIFNMFPMANPLVLYAVVQCAFMADIVLRVHSGHIEFLT 411
 F+SA A FLTGVTPE+NYMFPMFAA+PLVLYAVVQ AFA AD+++LVHSGHIEFLT
 Sbjct: 63 FLSSAAVFLTGVTPEIFNMFPMFAALPLVLYAVVQGLAFASADLHLRVHSGHIEFLT 122

Query: 412 RVFMGIKAGLGGDIFNFVVTLLFAVLVYFIAMPMIKFNLATAGRNVDNEEVNAPS 471
 + FM IKAGL DI NF+ V+++F V MYFI NPMIKFNLAT+GRNVD + D +
 Sbjct: 123 KTFPAIKAGLADIVNFVIVSVFVGMVYFTFMKIKFNLATSGRNVDYDGD-DSDE 181

Query: 472 TAS-----GSADANSQVQVINLLGGRNIEVDACMTRLRVTYKQNSVGSBAAMKAGA 527
 TAS G+A+ANSQ+V+ +INLLGG+NI DVDACTMLR+TV D VG EAAMKAGA
 Sbjct: 182 TASNAGTANANSQIVKIINLLGGKNISVDACMTRLRITVVDVAKVGSBAAMKAGA 241

Query: 528 MGLVLKGNQVQVYGPKADVLKSDIQLDLSGTVIPVDLETGPVAAAPVTYKGITEE 587
 MGL+KGNQVQ+YGPKADVLKSDIQLDLSG IP D+ + A V ++KG+TEE
 Sbjct: 242 MGLVYKGNQVQVYGPKADVLKSDIQLDLSGVDIPKTDVTAPEEKTADV-SFKVTEE 300

Query: 588 IVSVAHQVVALDVVDKVPVFSQNMGGDFAVEPDTGNITYVPVSGVTYSVFPTKHAFLGLT 647
 +VA+GVQ + V DVPVFSQNMGGDFAVEP +GNITYVPVSGVTYSVFPTKHA FLGLT
 Sbjct: 301 VAIVADGVLPITQVEDVPVFSQNMGGDFAVEPDTGNITYSPVGLTYSVFPTKHAFLGLT 360

Query: 648 ESGLVLIHIGLITVALDQGPPEVKISSGQKVVAGDIADVADLRAKKA 696
 +GLEVLIVH+GLITVAL+G PP K+ GQ+V GDL +VADLRAKKA
 Sbjct: 361 DGLGLVLIHIGLITVALNGAPFSAKVKDGRVALDGLLADLRAKKA 409

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1667> which encodes the amino acid sequence <SEQ ID 1668>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -13.43	Transmembrane	186 - 202 (181 - 213)
INTEGRAL	Likelihood = -6.79	Transmembrane	419 - 435 (412 - 442)
INTEGRAL	Likelihood = -5.52	Transmembrane	61 - 77 (57 - 82)

-618-

INTEGRAL Likelihood = -3.56 Transmembrane 363 - 379 (363 - 381)
 INTEGRAL Likelihood = -1.97 Transmembrane 143 - 159 (142 - 160)
 INTEGRAL Likelihood = -0.16 Transmembrane 343 - 359 (343 - 359)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.6371 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AD00291 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 288/407 (70%), Positives = 331/407 (80%), Gaps = 2/407 (0%)

15 Query: 286 DLVHLGSD-ASAYSHLMSVTPARFKVQGMIGARTLNGVALAMYRVADKKHTYBM 344
 DL+HLKG+ S Y HL+ SVTPARFKVQGMIG+G LMG+ LAMRVAD DKK YK M
 Sbjct: 3 DLHLKGAGHMSQYHLLTSVTPARFKVQGMIGSSGILMGLTAMYRVDFDKKHTYBM 62

Query: 345 FLSAAAVFLTGVTPELEYLPMFAVFLYVVA+VQGSFAMADLVLRVHSPONIE 404
 F+SAA AVLTLGVTPELEY+FMFA+PLX+VVA+VQ+FA ADL+L+RVHSPONIE LT
 20 Sbjct: 63 FLSAAAVFLTGVTPELEYLPMFAVFLYVVA+VQGLAFASADLHLRVHSPONIEFT 122

Query: 405 RTPMAIKAGLMDVIRFVNVSVFVIMYFIADNMIIKIMHLATAGRLNYDA-DILGDN 463
 +TPMA+KAGL MD+NF+ VSV+V V MFI + MIK +LAT+GR GND D D
 Sbjct: 123 KTPMAIKAGLMDIVIRFVNVSVFVIMYFIADNMIIKIMHLATAGRLNYDA-DILGDN 182

25 Query: 464 TQTRPTQVADNSQVQIVNLLGGAGNIDVDACNTRLRITVTDVAKVGEDEAAWKAGAM 523
 A++NSQ+V+I+MLGG NI DVDA+CNTRLR+TV D AKVG E WKAGAA
 Sbjct: 183 ASNSNAGTANANSQIVKIINLLGGAGNIDVDACNTRLRITVTDVAKVGEDEAAWKAGAM 242

30 Query: 524 GLIVKKGAVQVQPKADILKSDIQLLDSGALIEPVNMSQLTSKPTAKDFKHVIRTVL 583
 GLI KKGAVQVQPKAD+LKSDIQLLDSG IP+ +++ T FK VTR+V
 Sbjct: 243 GLIVKKGAVQVQPKADILKSDIQLLDSGVDIPKTVTPASEDKIADVSPKGVITREVA 302

35 Query: 584 SVADGVLPIITGVKDKVFAKMGDGFAVEPETHGNIYAPVAGLATSVPFTHAGLITDN 643
 +VADG VLPI T V D VF+ KMGDGFAVEP +GNIY+PVAGLATSVPFTHAGLITDN
 Sbjct: 303 TVADGQVLPIITGVHDFVFSQKMGDGFAVEPETHGNIYAPVAGLATSVPFTHAGLITDN 362

Query: 644 GLEVLHVGLDITVALNGVFFSVKVSQGRVHAGDILLVADLAIAKSA 690
 GLEVLHVGLDITVALNG PFS KV +GQRV GDLL+VADL AIKSA
 40 Sbjct: 363 GLEVLHVGLDITVALNGAFPSAKVKIGQRVALGDILLVADLAIAKSA 409

An alignment of the GAS and GBS proteins is shown below:

Identities = 517/731 (70%), Positives = 606/731 (82%), Gaps = 7/731 (0%)

45 Query: 6 MKNNVQLSPFEWQKFGKALMVIAVWPAAGLWISIGNSISLDPSNVLLGRANVIAQ 67
 MK + KQLF FEWQKFGK L MVIAVWPAAGLW+SIGNSI +++ + L + M+IAQ
 Sbjct: 1 MKTSFKQLFRFEWQKFGKALMVIAVWPAAGLWISIGNSIPWIMHIDAFSLIGNHIAQ 60

50 Query: 68 IGMVGICNHLILFALAIIGSNWAKERAGGAFASGLIFILINLTGNPFGVKTMLADSKAT 127
 IGM VI MLH+LFAIAIGSNWAKERAGGAF+GL+P+LIN ITG P+GV + MLAD
 Sbjct: 61 IGMVAVINLHLLFALAIIGSNWAKERAGGAFASGLIFILINLTGNPFGVKSSTMLADPEAK 120

Query: 128 VQIVFGATIRVSDYFVNVLGQPALNGVFFVIGSGVFGATAPHKYNYNRKLPDALTFPFG 187
 + ++ G + V DYP +VL PALN GVFFVIG+GFVGAT+HKYNYNRKLP+ LPFPG
 55 Sbjct: 121 ITSLLGTQIVKDYFTSVLRSFALNIGVFFVIGIAGVFGATAYNKYNYNRKLPDLVITFPFG 180

Query: 188 KRFVPFVVIYRSVIVAILLSVFWPVPVQSGINGFGKWIASSQDSAPILAPPVYGLTERILL 247
 KRFVPFVVI RS+ VALIL V WPV+QSGIN FG WIASSQDSAPILAPP+YGLTERILL
 Sbjct: 181 KRFVPFVVIILRSIFVALILVWVPVQSGINGFGKWIASSQDSAPILAPPVYGLTERILL 240

60 Query: 248 PFGLEHMLTIPNNTYQLGTTYTLVATGKAQVLGQDPLMLAWGLDILNKGSSSYVHH 307
 PFGLEHMLTIPNNTY LQTTY V+TGA G +V QGDPLMLAW DL++LKGSS+8 Y H
 Sbjct: 241 PFGLEHMLTIPNNTYALQTTYXWMTGAAGTKVFGQDPLMLAWTDLVHLKSSDASAYSH 300

65 Query: 308 LLTSVTPARFKVQGMIGASGILMGLSYAMYNVDKDKKIKYKSMPTISAAAPFTLTGVTPE 367
 L+ SVTPARFKVQGMIG+G LMG++ AMYRNVD DKK YK MPTISAAA FLTGTPE

-619-

Sbjct: 301 LMSVTEPARFKVGMIGATCTLMGVVALMYRNVDDAKGHTYRMFLSAAAVFLTGVTEP 360

Query: 368 IEXMFMEFAMPLYLVAVVQGCAPAMADIVNLRVHSFGNIETLTVPMGIKAGLGSDIFN 427
+EY+MFAMPLY+VTA+VQG +FAMAD+VNLRVHSFGNIETLTVPM+KAGLG+D+N

5 Sbjct: 361 LEVLFMFAMPLYIVYALVQGSFAMADLVNLRVHSFGNIETLTVPM+KAGLG+D+N 420

Query: 428 FVWVILLFAVLMYFIANFMIKKMLATAGRGNYDNSEVD--NAPSTAGSADANSGVQ 485
FVWV+LFAV+MYFLA+ MIKK+LATAGR GNYD + + N + + AD+NSGVQ

10 Sbjct: 421 FVWVSVLFAVIMYFIADMMIKKMLATAGRLGNYDADILGDRNTQTRPTQVADNSGVQ 480

Query: 486 VINLLGGRDNIEDVDACMTRLEVTVKDGSVGSAAWKKAGAMGLVAKGNGVQAIYGEKA 545
+NLLGGS NI+DVDACMTRLEVTVKD VG+E WKAGAGL+KMGVQQA+YGEKA

15 Sbjct: 481 IVNLLGGAGNIDVDACMTRLEVTVKDPAKVGARDWKAGAGLIGLIQKRGVQAVYGEKA 540

Query: 546 DVLKSDIQDLLSGTIVPIVLE--TGQPVAAAPVTYKQITTEIVSVANGQVRLDVK 603
D+LKSDIQDLLSG +IP V++ T+D P +K +TE++SVA+G V + VK

20 Sbjct: 541 DILKSDIQDLLSGALIPVNMSSQLTSKP--TRAKDPRHVTEDVLSVADGMVLFTITGVK 597

Query: 604 DPVTSQRMGDSGFAVEPTDGNIIYVPVSTVTSVFTTHARGLLTESGLEVLWHIGLDTVA 663
D V+ R+MGDSGFAVEPT GNIIY PV+G VTSVFTTHARGLLT+GLEVLWH+GLDTVA

25 Sbjct: 598 DQVFAARKMGDSGFAVEPTHGNIYAPVAGLTVSVFTTHARGLLTDNGLEVLVHVGGLDTVA 657

Query: 664 LDGQPFVKISSGQKVAGDLAVVADLEAKAAGKETSIIIVFTNVSIDKTVKLEKSGPQ 723
L+G PF VK+S GQ+V AGDL VVADL AIK+A +ET +++ FIN ++I+ V L G Q

30 Sbjct: 658 LAGVPPSVKVSQGRVHAGDLLVADLEAKAISKARETIIIVAFNTWTRIQDVTLSLGAQ 717

Query: 724 IAKTVAKVEL 734
AKT VA VEL

30 Sbjct: 718 PAKTKVATVEL 728

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 522

A DNA sequence (GBSx0560) was identified in *S. agalactiae* <SEQ ID 1669> which encodes the amino acid sequence <SEQ ID 1670>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2266 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 523

A DNA sequence (GBSx0561) was identified in *S. agalactiae* <SEQ ID 1671> which encodes the amino acid sequence <SEQ ID 1672>. This protein is predicted to be alkaline phosphatase synthesis sensor protein phor (hpkA). Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-13.96 Transmembrane 160 - 176 (148 - 183)
INTEGRAL Likelihood = -8.65 Transmembrane 20 - 36 (13 - 41)

-620-

----- Final Results -----

5 bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8595> which encodes amino acid sequence <SEQ ID 8596> was also identified. Analysis of this protein sequence reveals the following:

10 Lipop: Possible site: -1 Crend: 6
 SRCFLG: 0
 MG: Length of UR: 26
 Peak Value of UR: 3.27
 Net Charge of CR: 3
 15 MG: Discrim Score: 14.63
 GVH: Signal Score (-7.5): -5.64
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 Amino Acid Composition: calculated from 1
 ALOM program count: 2 value: -13.96 threshold: 0.0
 20 INTEGRAL Likelihood = -13.96 Transmembrane 152 - 168 (140 - 175)
 INTEGRAL Likelihood = -8.65 Transmembrane 12 - 28 (5 - 33)
 PERIPHERAL Likelihood = 1.59 135
 modified ALOM score: 3.29
 icml HYPID: 7 CFP: 0.658
 25 *** Reasoning Step: 3

----- Final Results -----

30 bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS gene <SEQ ID 8593> and protein <SEQ ID 8594> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 6
 MG: Discrim Score: 14.63
 GVH: Signal Score (-7.5): -5.64
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 40 ALOM program count: 2 value: -13.96 threshold: 0.0
 INTEGRAL Likelihood = -13.96 Transmembrane 152 - 168 (140 - 175)
 INTEGRAL Likelihood = -8.65 Transmembrane 12 - 28 (5 - 33)
 PERIPHERAL Likelihood = 1.59 135
 modified ALOM score: 3.29
 45 *** Reasoning Step: 3

----- Final Results -----

50 bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

34.9/61.1% over 363aa
 55 Thermotoga maritima
 EGAD|131465| sensor histidine kinase HpkA Insert characterized
 GP|1575578|gb|AAC44437.1||U67196 histidine protein kinase Insert characterized
 GP|4982228|gb|AAD36721.1|AE001807_12|AE001807 sensor histidine kinase HpkA Insert
 characterized
 60 PIR|C72228|C72228 sensor histidine kinase HpkA - (strain MSB8) Insert characterized
 ORF0680(919 - 1977 of 2277)

XGAD|1331465|TM1654 (48 411 of 412) sensor histidine kinase HpkA (Thermotoga maritima)
 GP|1575578|gb|AAC44437.1||U67196 histidine protein kinase (Thermotoga maritima)
 GP|4982228|gb|AAD36721.1|AA001807.12|AA001807 sensor histidine kinase HpkA (Thermotoga
 maritima) PIR|C72228|C72228 sensor histidine kinase HpkA - Thermotoga maritima (strain
 MS8)
 %Match = 13.6
 %Identity = 34.8 %Similarity = 61.0
 Matches = 125 Mismatches = 134 Conservative Sub.s = 94

[illegible]

>GP:BA306875 GB:AP001517 two-component sensor histidine kinase
involved in phosphate regulation [Bacillus halodurans]
Identities = 176/589 (29%), Positives = 315/589 (52%), Gaps = 47/589 (7%)

50	Query: 9	MTKKIPFPTTSLASGLVLTVMIMG-----FLYNNPHNQIQARQQTALAS	56
	MTK +R L+ ++ VT+L+ G	+L N + ++E ++ +	
	Subject: 1	MTKPYRLVLA---VLVT+LVNAGGLVGIGIPKNNYLTITRLKKTQTLAASW	56
55	Query: 57	QGISE-EGDYFENLKTS-NVRITWVWNGKQVLYTQSDAKHNMHNAHRKEIKESGY	114
	+ F + E + + R + T + G V + + + D M + N + A + R E E + + G		
	Subject: 57	EAVLENEVCTLTETKISQKLDARVTIILVLTGVGSAAPMAGNNAARPEPE-LESGI	115
	Query: 115	GESTRWSMIL-TKESIYAQRNLN-NGTII-VRLSVAQQTIFYLLGLMISPLAIIILL	169
60	R+S T + E YA N N T I VRL + + + + + L + + + A		
	Subject: 116	--VRYSTVETRLGLPYAVIQNEANETIGYRLGLPIEAVNSVNTMLVLSPTAF	172
	Query: 170	ILSVLIARYIAKKVSEPLNNI-----DLDPLDLSNVEEITPLRLRLDSHQAKI	219
	+ + V + IA + + P + + D S + + E + L R + +		
65	Subject: 173	LVTVSVTYIAQNMIRIPSEPTVANKIABDQARTSEESRDQVRSINVLVAAL	232
	Query: 220	HQKLLKQKQKQFPTTISKTSGEMILLDQARTVSIINAEALKLPQIND-NHGRFAMEVS	278

-622-

Q +++ +T+I + G+IL++ + I IN +FQ + D N + +V
 Sbjct: 233 QUTKRHQVQKERIKTLIRNNGSGILILINFRGDISLHKYCHDIPQRIYDGLNLAQLYHDV1 292

Query: 279 RLITLAKDLIDQGLKGGKKKAN-----IGLENNHYKVLNRPFT-DNNRVTLGLVLAFLVDV1 332
 + + ++ +K++ I +E H+ V P +N++ G+ ++ D+T+

5 Sbjct: 293 KHKEIKIVQDIFLTKRQRQVKLPFHLEVRHFDVHGAPIVRKNKGIKGLIALVPHDITE 352

Query: 333 QLQMEQLQRREFTANVSHLKTPLHVISQYSELLANQMVNEHV-PQFAKIHKESRLVK 391
 ++EQ++++F ANVSHLKT+ I G+E L + + +E++ CF I KESERL

10 Sbjct: 353 LKLEQVQRKDFVANVSHLKTPTVTSIKGPTITLLDGMIDEQLRQDFLHIWKESERLQS 412

Query: 392 LVEDIINLSHLDEQE-KLPQETPVNLYDLTPQVLEGLQAKDKKHQINFMGERAI-LRGN 449
 L+ D++ L\$ +++ +L + N++ + +V+ L+ KA++K I I+ + E + L G+

15 Sbjct: 413 LHDLELESLKIEQYIPQLNWOQTNLPAVVSFVMTLLKGAEEKIDISLSAEGSFOLBDD 472

Query: 450 FVLINSLVYNLCINAITYNHEKQGVNVTLRNSPDTITLSESDTGLGIAEKDKKRIFERP 509
 P L + NL +NAITY G++++ LK+ D + EV+DTG+GI E + RIFERP

Sbjct: 473 PERLAQLAINLVNNAITYTNSGRIDIALKDHGDVVEPVDNTGIGRESEIPRIFERP 532

20 Query: 510 RVDKRSKIGCTGGLGLSIVKALDPHNGSIKVDSHLGGQGTMTVLLHK 558
 RVD++RS+ GTGLGL+IVK ++ H G I V+S G+GTT T+ H+

Sbjct: 533 RVDARSRNSGCTGLGLAIVKLVKHAHQKILVSESPGKGTTFITQIFR 581

There is also homology to SEQ ID 1178.

25 SEQ ID 8594 (GBS340) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 10; MW 86kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 7; MW 61.5kDa) and in Figure 77 (lane 10; MW 62kDa).

Purified GBS340-GST is shown in Figure 223, lane 2; purified GBS340-His is shown in Fig. 191, lane 9.

30 The purified GBS340-GST fusion product was used to immunise mice. The resulting antiserum was used for Western blot (Figure 254A), FACS (Figure 254B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 524

A DNA sequence (GBSx0562) was identified in *S. agalactiae* <SEQ ID 1673> which encodes the amino acid sequence <SEQ ID 1674>. This protein is predicted to be phosphate regulon transcriptional regulatory protein phoB (phoB). Analysis of this protein sequence reveals the following:

40 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.2617 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10203> which encodes amino acid sequence <SEQ ID 10204> was also identified.

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC73502 GB:AE000146 positive response regulator for pho

-623-

regulon, sensor is PhoR (or CreC) [Escherichia coli K12]
Identities = 98/224 (43%), Positives = 138/224 (60%), Gaps = 2/224 (0%)

5 Query: 2 IYCVEDDADIREMMLYTLQAGFKAGQFSSSELPFWAIEQKVPDLILLDMLPGDGLTI 61
I VED+A IREM+ + L+ GF+ + + E PDLILLD MLFG G+
Sbjct: 5 ILVVEDEAIPIRMVCFVLQNGFQVFEARDYDSAVNQLNEPWFOLLDDMLFGGSGIQF 64

10 Query: 62 LERLRIGHCTEMIPVIMTTAKSGEYDKVGLDLGADNLYKPGMMISRIKAVLRGR 121
++ L+R+ T IF+M TR+G E D+VGL+ GADY+ KPF R++RIKAV+RR
Sbjct: 65 IKGHLKRESMTRDIPVWMLTARGEEDRVRSLETGADDTYTKPFSPIGLVARIKNMMRIS 124

15 Query: 122 QVDSKAHIIIGNLEIDPTNYVWVKRGTCKLHLTKGEPELLVLPFRNPNRVPTRQLLDKVM 181
+ + I + L +DPT+ V G E + + E+ALL F +P RV++R++LL+ VW
Sbjct: 125 PMAVEEVIMQGLSDPTSHRVNAGEEPLKMGKPTFEKLLHFFMTHPERVSRBQLLAHW 184

Query: 182 GEQFLGEHTKTVFVNHGTLATKLGEQDY--LIATVRGVGYRLEER 223
G E RIVVHI LR L G+ ++ TVRG GYR R
Sbjct: 185 GTNVYVEDRTVVDHIRLRKALEPGGHDMMVQTVRGTYGRFSTR 228

20 There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 525

25 A DNA sequence (GBSx0563) was identified in *S. agalactiae* <SEQ ID 1675> which encodes the amino acid sequence <SEQ ID 1676>. This protein is predicted to be phosphate transport system regulatory protein (phoU). Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1188 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG08750 GB:AE004948 phosphate uptake regulatory protein PhoU
[Pseudomonas aeruginosa]
Identities = 66/213 (30%), Positives = 119/213 (54%), Gaps = 4/213 (1%)

40 Query: 2 IGRSFASQLNDLNKEIIFMGALCEDIIKSLGALTNSNDVYLDIDSETYHKIQMERDIE 61
I +F ++L D+ ++ M3 L E + ++ AL +++ + E +I QMER+I+
Sbjct: 11 ISQFNALBEDVRSHLLAMGLHVEKQVNDVNALIDSLGACQVREIDQYQMERID 70

45 Query: 62 ERCLKLLLRQOPVAKLRRISSALKRWVDMKRIQAQAYEIAIVSLGHIIQSGSERD-- 119
E C++L R+Q P A DLR I S K V D++RIG +A ++A + + S R
Sbjct: 71 BECVRIARROPAPASDLRLISISKSVIDLERIGDFASKVARRAI--QLCEEGSPRGV 128

50 Query: 120 QLSMSENNSVIMLTKEIDAFIYDNEEQAHQVIEQDRIVNQEPDTIKKQWLVSQVDVG 179
++ + + V M+ ++DAF + + A V + D+TV++E+ T ++L Y
Sbjct: 129 EVRHISQVQKMQVQELDAFADPADLALSVAQYDKTVQREYKTAELVYTMEDPRAI 188

Query: 180 EYPIDVIMIAKYLERIGDMITVNIKAVLFSITG 212
++++ + LERIGH NIA+ V++ + G
Sbjct: 189 SRVLNINALESLERIGDMARNIAELVIVYLRG 221

55

There is also homology to SEQ ID 1678.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 526

A DNA sequence (GBSx0564) was identified in *S.agalactiae* <SEQ ID 1679> which encodes the amino acid sequence <SEQ ID 1680>. This protein is predicted to be ATP-binding cassette protein PstB (pstB-2). Analysis of this protein sequence reveals the following:

```

5      Possible site: 52
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2432(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10205> which encodes amino acid sequence <SEQ ID 10206> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

      >GP:AAD22041 GB:AF118229 ATP-binding cassette protein PstB
      [Streptococcus pneumoniae]
      Identities = 166/245 (67%), Positives = 211/245 (85%), Gaps = 1/245 (0%)

20      Query: 10  INNLDLYYGEFHAKDVNLDIEEKSTAFIGPSGCGKSTLLKSNMNDLVKNCKITGDI 69
      +LDL+YG+F ALK++++ + E++ITA IGPSGCGKST LK++NMNDLV +C I G +
      Sbjct: 6  VRHLDLFYGDGFQALKNIISIQLEPQITALIGPSGCGKSTFLKTLNRMNDLVPSCHIEGQV 65

25      Query: 70  TLBEGEDVYR-QLDINLQKRVGMVFPQKPNFPMISIYDNVAFGPRTHGHISKAELDIVER 128
      L+ +D+Y + ++NQLRK+VGMVFPQ+PNFF MSIIDNV+A+GPRTHGI K +LD +VE+
      Sbjct: 66  LLDEQDIYSSKFNLNQLRKNVGMVFPQPNFPMISIYDNVAYGPRTHGIRDKKQLDALVER 125

30      Query: 129  SLNGAALMEVDIKRLEKSAIGM/GSGGQQRLCTARALAEIPVLMDPTSLALCPISTAKI 186
      SLK AA+W+EVKD L KSA+ +SGGQQRLCTARALAEPC+LLMDEPTSLALCPIST KI
      Sbjct: 126  SLNGAALMEEVKDLKRSAMSLSGGQQRLCTARALAEPCILLMDEPTSLALCPISTAKI 185

35      Query: 189  EELVIQLKKNTTIVITHNMQQAVRISDKTAFFLMGHVEVYINKTSOLFSLPQERTENYI 248
      E+L+ QLEK+YTI+IVTHNMQQA RISDKTAF L GE+ E+ T +F+ P+D+RIE+YI
      Sbjct: 186  EDLIQLKKNDYTIIVITHNMQQASRISDKTAFFLMGHVEVYINKTSOLFSLPQERTEDYI 245

      Query: 249  TGRFG 253
      +GRFG
      Sbjct: 246  SGRFG 250

```

40 There is also homology to SEQ ID 1682.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 527

45 A DNA sequence (GBSx0565) was identified in *S.agalactiae* <SEQ ID 1683> which encodes the amino acid sequence <SEQ ID 1684>. This protein is predicted to be transmembrane protein PstA (pstA-2). Analysis of this protein sequence reveals the following:

```

      Possible site: 38
      >>> Seems to have a cleavable N-term signal seq.

50      INTEGRAL  Likelihood = -13.11  Transmembrane  265 - 281 ( 255 - 286)
      INTEGRAL  Likelihood = -8.81   Transmembrane  79 - 95 ( 68 - 100)
      INTEGRAL  Likelihood = -4.78   Transmembrane  195 - 211 ( 192 - 213)
      INTEGRAL  Likelihood = -4.67   Transmembrane  147 - 163 ( 143 - 164)
      INTEGRAL  Likelihood = -2.92   Transmembrane  122 - 138 ( 120 - 138)
      INTEGRAL  Likelihood = -0.90   Transmembrane  40 - 56 ( 39 - 56)

```

-625-

----- Final Results -----

bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA02040 GB:AF116229 transmembrane protein PstA [Streptococcus pneumoniae]
 Identities = 135/263 (51%), Positives = 203/263 (76%)

10 Query: 23 FFLFAIVYLGAISLPATIAFVVIYILVKGLEPHVNTGLFAMTYNTQNVSLPAMFIMTIFII 82
 + L +VY + L+F ++ ++ +IL+KGLPH+ LF+WTY ++N+SL+PA I+T++ ++
 Sbjct: 4 YLLKLVYCFPSALTFGSLFLIIGFILINGLPHLSLPSWTTTSENISLMPAISTVILV 63

15 Query: 83 ALTLFAVPLGIGGSIYLTETARRDNFYKLIIRVATETLAGIPSIITGLFGALFFVKYTH 142
 LL A+P+GI YL EY ++D+ +KI+R+A++TL+GIPSI++GLFG LFFV +
 Sbjct: 64 FGALLALFTIGIFAGFYLVETTKKDSLCKIMRLASDTLGIPIIVGLFGMLFFVFLG 123

20 Query: 143 LGLSLISGSLTSLIMILFLIMRTTEALLSVFDSYREGAFALGAGKLRITFKIVLPSAMS 202
 SL+SG IT IM+L+I+R+TEALLSV DS R+ ++ LGAGKLRIT+P+IVLF AM
 Sbjct: 124 FQYSLSGIITSVIMVLPVLIIRSTEALLSVSDSMRQASVGLGAGKLRITVPRIVLFPVMP 183

25 Query: 203 GIFAGIILAVGRIGBSAALIFTAGTVAKVAHSVFSSSRLTAVHMYAISGBGLYVDQTYA 262
 GI AG+ILA+GRI+GE+AL++T GT S+ SS R+LA+HMY +S EGL+V++ YA
 Sbjct: 184 GILAGVILATGRIVGETALMYLTSTNTFPSSIMSSGRSLALRWMLSEGLHVNAYTA 243

Query: 263 TAVILLLLVIIWVNFVSGLVAKRL 285
 T VIL++ V+++N +S L++++L
 Sbjct: 244 TGVIILITVLMINTLSSLKRL 266

30 There is also homology to SEQ ID 1686.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 528

A DNA sequence (GBSx0566) was identified in *S.agalactiae* <SEQ ID 1687> which encodes the amino acid sequence <SEQ ID 1688>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2687(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 529

A DNA sequence (GBSx0567) was identified in *S.agalactiae* <SEQ ID 1689> which encodes the amino acid sequence <SEQ ID 1690>. This protein is predicted to be transmembrane protein PstC (pstC-2). Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

-626-

```

INTEGRAL Likelihood = -10.67 Transmembrane 256 - 272 ( 251 - 279)
INTEGRAL Likelihood = -8.86 Transmembrane 141 - 157 ( 133 - 162)
INTEGRAL Likelihood = -4.99 Transmembrane 111 - 127 ( 109 - 132)
INTEGRAL Likelihood = -4.30 Transmembrane 76 - 92 ( 72 - 95)
5 INTEGRAL Likelihood = -1.86 Transmembrane 25 - 41 ( 24 - 42)
INTEGRAL Likelihood = -1.33 Transmembrane 59 - 75 ( 59 - 75)
INTEGRAL Likelihood = -0.27 Transmembrane 203 - 219 ( 202 - 219)

----- Final Results -----
10 bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:AAD22039 GB:AF118229 transmembrane protein PstC [Streptococcus pneumoniae]
Identities = 162/266 (60%), Positives = 212/266 (78%), Gaps = 3/266 (1%)

Query: 15 ITACVSVISAILICLFPSGLRAITKIGWGNFIFGKVHPSN--NIPGIFPMIVGSLIV 72
++A V+V++ +LIC F+FS+GLP I G+ F+ G W P+N +GI PMIVGSL +
20 Sbjct: 1 MSATVAVVAILLICFFIFSNGLFFIANYGFARFLLGSDWSPTNIPASVIGILPMIVGSLAI 60

Query: 73 TAGALLGGPIGILTAVPMAYFCPENITYKPLKSAINLMAGIPSVVYGGFGLVIVVFMIRQ 132
T GR++G P GILT+VFM Y+CP+Y LKSAINLMA IPS+VYGGFGL ++VF IR
25 Sbjct: 61 TLGALIVGVPTGILTSVFMVYCPKPYGVFLKSAINLMAAIPSVYGGFGLQLLVFMIRS 120

Query: 133 YIGGFGMGVLAASILLGIMILFTTIVSISBSLRAVPESYIEGGIALGASHERSVFFAVLP 192
++G GM VL AS+LLGIMILFTI+S+SES++R VP++YI G +ALGASHERE++F +LP
30 Sbjct: 121 FLGN-GMSVLTASLLGIMILFTIISLSBSAIRTVPKITYGSLALGASHERSIFSVLP 179

Query: 193 AARSGILASVVLGIGRAIGETMAVIMVAGVAVLPQSLTSGVRTLTTHIVMBGYSGLH 252
AA+ GIL++V+LGIGRA+GETMAVI+VAGNQ ++F L SG RTLTTHIV+EM Y+SG H
35 Sbjct: 180 AARSGILSAVLGIGRAIGETMAVILVAGVAVLPITPSGLFSGRTLTTHIVLEMAVAGSQ 239

Query: 253 RQALIGTAVVLFIPIFILMINISPSALQ 278
R+ALI T+ VLF IL+IN F+ L+
40 Sbjct: 240 REALIATSAVLFFILLINAYFAYLK 265

```

There is also homology to SEQ ID 1692.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 530

A DNA sequence (GBSx0568) was identified in *S.galactiae* <SEQ ID 1693> which encodes the amino acid sequence <SEQ ID 1694>. This protein is predicted to be probable hemolysin precursor (pstS). Analysis of this protein sequence reveals the following:

```

45 Possible site: 34
>>> May be a lipoprotein

----- Final Results -----
50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

55 >GP:AAD22038 GB:AF118229 phosphate binding protein PstS
[Streptococcus pneumoniae]
Identities = 134/295 (45%), Positives = 185/295 (62%), Gaps = 9/295 (3%)

Query: 1 MKKHOMLSLLAVSGLMGIGILAGCSNDSSSSSK--GTINIVSREBGSQTRGAFIELFGI 57
NK KML+L A+ GL G G+A C N S+S + GTI ++SRE GSGTRGAF E+ GI

```


-627-

Sbjct: 1 MKFKIMLTAAI-GLSGFLVA-CGNQSAASKQASGTIEVISRENGSGTRGAPTEITGI 58
 Query: 58 ESNKKGKEVDHSDATVINSTSVMLTTVSKDPSAIGYSSGLASSVKVLKIDGNAT 117
 K4 +K4D+T+ A + NST +L+ V + +AIGY SLGL SVK L+IDG A+
 5 Sbjct: 59 LKKIGD-KKTONTAKTAVIQNSTEGVLSAVQGNANAGIYISLGLSKTVKALEIDGVKAS 117
 Query: 118 VKDIKSSSYKISRPFNITVKSGKEKATKDFIDYLSKDDQAVVEKNYIPL-DNAKAYQ 176
 + G Y + RPFNIV K +DFI +I SK QQ VV N +I - Y
 10 Sbjct: 118 RDTVLDEGYPLQPFNIVWSSNLSK-LGQDFISFHSKQQVVDNKPFEAKTETETET 176
 Query: 177 AKVSSGKVVIAGSSSVTPVMEKIKRAYHKVNAKVVDVEIQSDSSSTGITSADGSADIGMA 236
 ++ SK+ + GS+SV+ +MEK+ RAY K N +V ++I + SS GIT+ + +ADIGM
 Sbjct: 177 SGHLSGKLSVVGSTSVSSLMELAKAYKKNPEVTIDITNSGAGITAVKEKTADIGMV 236
 15 Query: 237 SRELKTESSEGVKATVIATDGIADVNNKKNVNDLSKQVKDIFTGHTTWSDL 291
 SREL E K + IA DGIADVNN NK + +S ++ D+P+CK T+W +
 Sbjct: 237 SREL-TPEEGKSLTHDAIALDGIADVNNNDKASQVSMALADVFSKGLTIWOKI 290

There is also homology to SEQ ID 1696.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8597> and protein <SEQ ID 8598> were also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: 23 Crend: 4
 MG: Discrim Score: 7.91
 GVH: Signal Score (-7.5): -3.72
 Possible site: 34
 >>> May be a lipoprotein
 30 ALOM program count: 0 value: 2.44 threshold: 0.0
 PERIPHERAL Likelihood = 2.44 248
 modified ALOM score: -0.99
 *** Reasoning Step: 3
 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 40 SEQ ID 1694 (GBS24) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 9; MW 33kDa).

GBS24-His was purified as shown in Figure 194, lane 10.

Example 531

- 45 A DNA sequence (GBSx0569) was identified in *S.agalactiae* <SEQ ID 1697> which encodes the amino acid sequence <SEQ ID 1698>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1725 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 532

A DNA sequence (GBSx0570) was identified in *S. agalactiae* <SEQ ID 1699> which encodes the amino acid sequence <SEQ ID 1700>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2741 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:BA05069 GB:AP001511 unknown conserved protein [Bacillus halodurans]
Identities = 119/250 (47%), Positives = 149/250 (59%), Gaps = 9/250 (3%)

Query: 1 MQQYFVNGE--AGAYVTIEDKDTIKHMFVNRITEDDQVVLVFDDAIKRLAKVVDSSAHR 58
MQ+YFV E YVTI D +KH+ VNR+T D+ L+ D R + A+
20 Sbjct: 1 MQRYPFVKEQVTDYTYVTITGDD-VKHIIKVMRMTIGDE--LICSDGSHGRTVRCETEKAND 57

Query: 59 FQIL----EELDNNVEMPVQVTIASGFPKGDKLDPVTKATELGAAIHWGFPADWSVVK 114
++L E L N E+P++VTIA PKGDKLD++ QK TELGA A W F A S+VKN
Sbjct: 58 SEVLARVIEFLIPNTELPPIRVITIAQALPKGDKLDYIVQKGTGGAQAPWPFASRSIVKN 117

25 Query: 115 DGKILAKKDKLAKIALGAAEQSGKNRILPQVRLFEKADFOAELAGFKFIAYEESAKE 174
D KK KK ++L KIA AAEQS R R+P + E++GF K +AYE AKE
Sbjct: 116 DEKKGRKCTELMKIANKAAEQSYRERIPISETFLPALSFKLLQKISGFTKTVIYEEAKE 177

30 Query: 175 GELSALAQWLQTVKAGDKLLFIQPEGGISPKETAAFEVGAIVGLGIPRIIMRTETAPLY 234
G L A L + GD LL I GPEGG + +EI A + G GIGPRI+RTETIA LY
Sbjct: 178 GRIMTFPAAICNLNHHGDELLVIQPEGGFTTEIDAIQKAGGAPAGLGPRIILRTETASLY 237

Query: 235 ALSVISYSAE 244
AL+ ISY E
35 Sbjct: 238 ALAASISYFE 247

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1701> which encodes the amino acid sequence <SEQ ID 1702>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2274 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 173/245 (70%), Positives = 202/245 (81%)

50 Query: 1 MQQYFVNGEAGAYVTIEDKDTIKHMFVNRITEDDQVVLVFDDAIKRLAKVVDSSAHRFQ 60
MQQYF+ G+A VTI DKDTIKHMF VMRL ++ +VVLVFD +K LAKV +S AH +
Sbjct: 1 MQQYPIKGKAEKIVTITDKDTIKHMFQVMRLADEAEVVLVFDGQVYLAKVNSMAHLE 60

55 Query: 61 ILEELDNNVEMPVQVTIASGFPKGDKLDPVTKATELGAAIHWGFPADWSVVKDQKGLA 120
I+E L + V+P+V+VTIASGFPKGDKLD + QK TELGA+A+WG+PADWSVVKDQKGLA
Sbjct: 61 ILENLPQVELPVKVTIASGFPKGDKLDITLAKQVTELGASALWGYPADWSVVKDQKGLA 120

Query: 121 KKEDKLAKIALGAABQSGKNRILPQVRLFEKADFOAELAGFKFIAYEESAKEGELSAL 180

-629-

KKEDFLAKI LGAAEQSKRRNP+P+V LFE KA+P L+ FD IFIAYEE+AK G+L+ L
 5 Subjct: 121 KKEDFLAKI VLGAAEQSKRRNP+P+V LFE KA+P L+ FD IFIAYEE+AK G+L+ L 180

Query: 161 AQNLQTVKAGDKLLFIQPGEGGISPKSIAAFEEVGAIKVGLSPRIMRTETAPLYALSVIS 240
 A+ ++ VK G K+LFIQPGEGGISP EI PE AIKVGSLSPRIMRTETAPLYALS + D
 5 Subjct: 161 AREVKEVKPGAKILFIQPGEGGISPRTITQFRAASAIVKVGSLSPRIMRTETAPLYALSALS 240

Query: 241 YSARL 245
 Y+ EL
 10 Subjct: 241 YSARL 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 533

- 15 A DNA sequence (GBSx0571) was identified in *S. agalactiae* <SEQ ID 1703> which encodes the amino acid sequence <SEQ ID 1704>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 20 INTEGRAL Likelihood = -2.28 Transmembrane 238 - 254 { 237 - 254 }
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1914 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA02791 (GB:AB023064) orf35 [Listeria monocytogenes]
 Identities = 138/309 (44%), Positives = 193/309 (61%), Gaps = 5/309 (1%)

30 Query: 4 VNELTVHVNREAEVSNLLIETSGSQVAISDSADYLGQ-EDRFGELYP---EVEQSDMI 59
 W+E+ VH EA E V+N+L E G+ GV+I D AD+L + ED+FG+E+ Y E + D
 Subjct: 3 WSEVKEHTTNEAVEP+VANVLTGFAAGVSIEDVADFLREREDKFGELYALRREDYEDGV 62

35 Query: 60 AITAYYPDTLDEAVKADLADLANFSGFGLATGSGVNLDSQELVEEDWADNKKYTPAR 119
 I AY+ T + ++ L N F + G ++ +E+WA NKKY P +
 Subjct: 63 IIKAYFLKTEFVVRQIPBIRQFLKLNISTFDIPLGKQFVVDVDDSEWATNKKYTHPQ 122

Query: 120 ITHDITVPSWTDYRAGKGIKIMDPQMGRTGTHTPTKSLFALEQLRGGEVTVDVG 179
 IT +ITVPSW Y A E II++DQMGRTGTHTPT++ +AL L+ G+ VINDG
 40 Subjct: 123 ITURITVPSWESYTPSANEIILDQMGRTGTHTPTQLCIRALSNYLQGGEDVIDVG 182

Query: 180 TGSQVLSTASSLIGAKDIYALDLDVAVRVAGQENIDMPTGTENIHVAGDLLEKVVQO-EV 238
 TGSQVLSTAS+ LGAK I A DLD+A R A+ENI +N I V ++L+ + + V
 45 Subjct: 183 TGSQVLSTASAKIGAKSLATLDLDTAAEENITANKTEHIIIVKQNLQDINKTV 242

Query: 239 DIVVANILADILIHLDYAVRVKDEGYLINSGLISEKMDWKEASAKGAPFLTHMVQG 298
 D++VANILAAAA +D Y+ +K G I SGII +K +V E+ + AG +E CG
 Subjct: 243 DIVVANILAEVILLFPEDVYKALKPGVFIASGIIEDKAKVKEAKNAGLIIKKMEQOQ 302

50 Query: 299 EKQACVKK 307
 +N A + K+
 Subjct: 303 DWVAIIKKR 311

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1705> which encodes the amino acid
 55 sequence <SEQ ID 1706>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 60 INTEGRAL Likelihood = -4.57 Transmembrane 238 - 254 { 237 - 257 }
 ----- Final Results -----

-630-

bacterial membrane --- Certainty=0.2826 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:BAAB2791 GB:AB023064 orf35 [Listeria monocytogenes]

Identities = 139/309 (44%), Positives = 203/309 (64%), Gaps = 5/309 (1%)

10 Query: 4 WQEVTVVHRDAQEAHSVLIETGSGQVAIDSDADYIQK DRFGELYP---DVEQSDMI 59
 W EV VH +A E V++VL E G+ GV+I D AD++ ++ D+PGE+Y + D+
 Sbjct: 3 WSEVVEFTTNEAVEPFAVNLTEFGAAAGVSIREDVADFLREREDKFGRIYALRRREDYPEDGV 62

Query: 60 AITAYPSSITLADIATINEQLAELASFGLVGQVTVDSQELASEDWADNWKYYEPAR 119
 I AY+ +T + I I + L L++F + +G+ ++ +E+WA WKYY P + D+
 15 Sbjct: 63 IIKAYFLKTTETVQIPEIQELKLNSTFDIFLGKQFVNVVDDEBWAANKYYHPVQ 122

Query: 120 ITHDLTIVPSWTDYDASAGEKVIKLDPGMAGTGTHTPTTMSLFALEQILRGSETVIDVG 179
 IT +TIVPSW Y SA E +I+LDPGMAGTGTHTPT++ + AL L+ G+ VIDVG
 20 Sbjct: 123 ITCRITIVPSWSTYTSANEIIIELDPGMAGTGTHTPTQLCIRALSNYQGGDEVIDVG 182

Query: 180 TGGSVLSIASSLLGAKTIYAYDLDVAVRVAQNDINQGTINIHVAAGDLLKGVQ-BA 238
 TGGSVLSIAS+ LGAK+I A DLD++A R A++NI LN+ I V +LL+ +++
 Sbjct: 183 TGGSVLSIASAKLGAKSILATDLEIATRAEENITNKTETHTVKNQNLQDINKTNV 242

25 Query: 239 DIVVANILADILVLLTDAYRLVKKEGYLLSGIIEKLDMLVAEAFSAGFFLETHMVQG 298
 D++VANILA++LL +D Y+ +K G I SGII +K +V RA +AG +E QG
 Sbjct: 243 DIVVANILAEVILLPDEDVYKALPGGVFIASGITEDKAKVBEALNAGLIIEKMEQGS 302

Query: 299 EWNALVFKK 307
 +W A++ K+
 30 Sbjct: 303 DMVAIIISK 311

An alignment of the GAS and GBS proteins is shown below:

Identities = 259/317 (81%), Positives = 287/317 (89%)

35 Query: 1 MNTWNEITVHVNRAREAVSNLLIETGSGQVAISDSADYIQKDRFGELYPVEQSDMIA 60
 M TW E+TVHV+R++A+EAVS++LIETGSGQVAI+DSADY+QK+DRFGELYP+VEQSDMIA
 Sbjct: 1 METWQEVTVVHRDAQEAHSVLIETGSGQVAIDSDADYIQKDRFGELYPDVEQSDMIA 60

40 Query: 61 ITATYPTDLDIEAVKADLADRLANFBGFLATGSVNLDQSELVEEDWADNWKYYEPARI 120
 ITATYP + ++ + A + ++LA FGL G V +DSQEL REDWADNWKYYEPARI
 Sbjct: 61 ITATYPSSTNLADIATINEQLAELASFGLVGQVTVDSQELASEDWADNWKYYEPARI 120

45 Query: 121 THDLTIVPSWTDYDASAGEKVIKLDPGMAGTGTHTPTTMSLFALEQILRGSETVIDVG 180
 THDLTIVPSWTDY+ A AGEK+I K+DPGMAGTGTHTPTTMSLFALEQ+LRGSETVIDVG
 Sbjct: 121 THDLTIVPSWTDYDASAGEKVIKLDPGMAGTGTHTPTTMSLFALEQILRGSETVIDVG 180

Query: 181 GSGVLSIASSLLGAKDIYAYDLDVAVRVAQNDINQGTINIHVAAGDLLKGVQGEVDV 240
 GSGVLSIASSLLGAK IYAYDLDVAVRVAQ+ND+N GT+NIHVAAGDLLKGV QE DV
 50 Sbjct: 181 GSGVLSIASSLLGAKTIYAYDLDVAVRVAQNDINQGTINIHVAAGDLLKGVQGEADV 240

Query: 241 IVANILADILHLITDAYRLVKKEGYLLSGIIEKLDMLVAEAFSAGFFLETHMVQGEW 300
 IVANILADIL+ LITDAYRLVK EGYLI+SGIIEK DMV E+A AGFFLETHMVQGEW
 Sbjct: 241 IVANILADILVLLITDAYRLVKKEGYLLSGIIEKLDMLVAEAFSAGFFLETHMVQGEW 300

55 Query: 301 NACVFKKTDIDISGVIGS 317
 NA VFKKTDIDISGVIGS
 Sbjct: 301 NALVFKKTDIDISGVIGS 317

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 534

A DNA sequence (GBSx0572) was identified in *S.agalactiae* <SEQ ID 1707> which encodes the amino acid sequence <SEQ ID 1708>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4198 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 535

A DNA sequence (GBSx0573) was identified in *S.agalactiae* <SEQ ID 1709> which encodes the amino acid sequence <SEQ ID 1710>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0683 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
Identities = 87/246 (35%), Positives = 139/246 (56%), Gaps = 13/246 (5%)

Query: 4 VKEVSI LSGSVKRTLHYDYKIGLFPPTALSEAGVRLYDDEALIRLQEIILFRELEFFPKD 63
VK+V+ +SGVS+RTLHYD I L P+AL++NGVRLY D L RLQ+IL F+E+ F L +
Sbjct: 5 VKQVARI SGVSI RTLHYDNI ELLNPGALT DGVRLYSDADLERLQQLF FKEIGFRLD 64

Query: 64 IKYLLDQAEKERODLLAQQIKLEWKRSHLEQVITHAKR--LQEKGDVYN----FDVYN 117
IK +L+ +R+ L Q ++L K+ ++++I R L G + MN F +
Sbjct: 65 I KEMLDHPNFDKGAALQSQKEILMKKKQKQMDMKIQTIDRTLLSVDGSETMKNKRDLFAGLS 124

Query: 118 KTELEQLQA---ENKQKWSQTAA--YKRFPAKHASDDFAQISOBMAKIMVQGLQKTON 171
++E+ Q E++ +G+ A ++ ++DD+ I E I +
Sbjct: 125 MKDIERHQQTAYDEVKLYGKEIAEETKRTSAYSADNRTIMAEFDSYIKRLIARMKHG 184

Query: 172 VDSQVQMCVKRLQDYISNIFYTCINELLAGLQMYQSDRFSQSIDKAGAGTSEFVSQ 231
D +Q V +D+I Q Y CT +I GLG++Y +D+RF+ SI++ G G + F + +
Sbjct: 185 PDDARTQAAGVAFRDEICQVHYDCTLDIFRGLGKLVYITDERFTDSINQY--GSLAFLRE 243

Query: 232 AIAYYC 237
AI YC
Sbjct: 244 AIIYYC 249

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1711> which encodes the amino acid sequence <SEQ ID 1712>. Analysis of this protein sequence reveals the following:

Possible site: 48

-632-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.28 Transmembrane 146 - 162 (143 - 167)
 INTEGRAL Likelihood = -2.92 Transmembrane 172 - 188 (171 - 190)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 40/107 (37%), Positives = 69/107 (64%), Gaps = 6/107 (5%)

15 Query: 7 YSTGELANLAGVSIRTVQYDQRGILIPALTAGGRRLYDSDLEQLMCLFLRDLGFSI 66
 Y ++A ++GVSIRK+ +ID +L P+ALT G RLY+D+DL+L+ I F ++GF +
 Sbjct: 3 YQVQVASISIGVSIRTLRHYNTELLNPSALTADYRLYSADLEKLQQLFFKSGIFRL 62

Query: 67 EQIRKVLAEENAAQVLELLLVHDIATAKEDLAAKEQQVDIAVKILDR 113
 ++I+++L N + L + KE L K+Q+D ++ +DR

20 Sbjct: 63 DEIKEMLDHFNFDKGAAL-----QSKWELMKKKQRMDKIQITDR 103

An alignment of the GAS and GBS proteins is shown below:

Identities = 40/133 (30%), Positives = 71/133 (53%), Gaps = 6/133 (4%)

25 Query: 6 EVSILSGVSVRLTHHYDKIGLPPPTALSEAGVRLYDDEALIRLQETILFRLEFPLKDIK 65
 E++ L+GVST+RT+ +YD+ G+ PTAL+ G RLY D L +L+ I R+L F ++ I+
 Sbjct: 11 ELANLAGVSIRTVQYDQRGILIPALTAGGRRLYDSDLEQLMCLFLRDLGFSIEIR 70

30 Query: 66 YLL--EQAKEERQDLAQQIKL----LEWKRSHLEQVITHAKRLQEGKDDYMFVYNYKT 119
 +L E A + + LL I L K ++ + RL+++ ++F +
 Sbjct: 71 KVLAEENAAQVLELLLVHDIATAKEDLAAKEQQVDIAVKILDRIRKQDQPSLDLFMDISL 130

Query: 120 EELQLEAKEKRW 132
 ++ +A K +W

35 Sbjct: 131 SMKNQKQWIKLQW 143

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 536

40 A DNA sequence (GBSx0575) was identified in *Sagalactiae* <SEQ ID 1713> which encodes the amino acid sequence <SEQ ID 1714>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.06 Transmembrane 57 - 73 (57 - 73)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14586 GB:Z99117 yrkN [Bacillus subtilis]
 Identities = 38/136 (27%), Positives = 60/136 (43%), Gaps = 3/136 (2%)

55 Query: 2 ITLQKAEASDLSEKIIA-IQRASFKAVYEKYHDQYDFYVEQIRKILVERPDCFYHVL 60
 L+ L+ A SDL+ +Q A AV E + D D + ++ + P + +L
 Sbjct: 9 VILELAKESDLPEPKKILQEAFAIAVINTFCDCSGPIPSDNDVQ-ESFNAGAAVVYHL 67

Query: 61 VDETIVGFLRLVIKDEKRAWLGTAILPLQYQQQGYGSANMALLETXYKLTAKVGLCTIA 120

-633-

D VQ + I + L + P+Y QG G +A +E YF W+ T
 Sbjct: 68 QCGKQVGGAVVRINSQTHNSLDLFRVSPETHSGHGLSAWKAIEAQYFVFLWETVTPY 127
 Query: 121 QEKLMRSFY-EKCGYH 135
 EK ++FY KCG+H
 Sbjct: 128 FEKEHINFPYVKCGFH 143

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 537

A DNA sequence (GBSx0576) was identified in *S.agalactiae* <SEQ ID 1715> which encodes the amino acid sequence <SEQ ID 1716>. This protein is predicted to be Bacterial mufT protein. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2417 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA06568 GB:AE004742 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 57/131 (43%), Positives = 82/131 (62%)
 Query: 10 FSGAKIALFCEBKGILSLRDEPDLFYAGFMDLPGGGRDNETPLCLFREVDEELSLTL 69
 FSGAK+ALF ++ RD+ P +P+ G+MD PGGGR ETP EC RE++EE S+ L
 Sbjct: 7 FSGAKLALFYGDHLVYKDEKPGIIPFGYWDPPGGGRBGLLETPARCALRELREEFIRL 66
 Query: 70 TRNHLVDVKTYSGLKFDKLSVFMGHISQKEYDSIVLGDGQDYKLMSIDEFLSHKKVI 129
 I+H + Y + F+V + +E+++I GDEGQ ++LM +D +L+H +
 Sbjct: 67 EEPRIKQCRQYPSTSGSAPPAYFLVARLEDEFEFAIRFGDEGQYRLMEVDAYLAHAMAV 126
 Query: 130 POLQRLRDYL 140
 P LQ RL DYL
 Sbjct: 127 PYLQRLGDYL 137

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 538

A DNA sequence (GBSx0577) was identified in *S.agalactiae* <SEQ ID 1717> which encodes the amino acid sequence <SEQ ID 1718>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3299 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-634-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1719> which encodes the amino acid sequence <SEQ ID 1720>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5527(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 111/156 (71%), Positives = 128/156 (81%)

Query: 1  MAKFGPLSVLEERLDKHLQYDFAMDMDKKNHVEVTFLEAQNSSAIEIVDDQGETSSD 60
      MA +GFLSVLEEE+DKH QYD+AMDMDKKNH VEVTF+LEAQN AI+T+DD GE + +D
Sbjct: 1  MATYTGPLSVLEERMDKHFDYDAMDMDKKNHAVEVTFVLEAQNKEAKITDSDGEVTDOD 60

Query: 61  IVFEDYVLFFNFVVKSRFDAEDYLVITIPYEPKKGLSREFLAYFAETLNEVATEGLSDLND 120
      IVFEDYVLFFNP KS+FDA DYLVITIP++ KKG SREFLAYFA+ LN+VA EG SDLND
Sbjct: 61  IVFEDYVLFFNPAKSCFDAADYLVITIPDAKKGFSREFLAYFAQLNDVAIEGHSDLND 120

Query: 121 LTTDSIEREGLSDWTDATFENGRAELKETEFPYPRY 156
      L DDS +F L W+ AFE G+ L+E YPYRY
Sbjct: 121 LADDSKADFTLEWNAQAFEGGQGLEEAASYPYPRY 156
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 539

A DNA sequence (GBSx0578) was identified in *S.galactiae* <SEQ ID 1721> which encodes the amino acid sequence <SEQ ID 1722>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2846(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP: CAB51273 GB: AL096872 putative acetyltransferase [Streptomyces
coelicolor A3(2)]
Identities = 35/109 (32%), Positives = 62/109 (56%), Gaps = 1/109 (0%)

Query: 51  VAEVDDKIAGVLDGFPYYPFAGKNHATP-GILLAEFYQQQGLGKALLKALLTEAKAQGY 109
      VAE+D +G + G P + HV G+ +A +G G+G+AL++A + EA+ +G+
Sbjct: 56  VAEIDGAVVGTVKLGFPPTPLASITHVRQIRGLAVAGNARGHGVGRALVRAAVHARHEGF 115

Query: 110 IKIAMEVGNNSRAISLYQKYGPTEARITKAFFIENHYVDALIFAKDL 156
      +I + V+G+N+ A LY+ GF E + F + + YVD ++ + L
Sbjct: 116 RRITLRVLGHNTAARGLYESGPFVVEGVQPEEFELDGRYVDDVLNQQL 164
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1723> which encodes the amino acid sequence <SEQ ID 1724>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```


-635-

bacterial cytoplasm --- Certainty=0.0229 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 34/108 (31%), Positives = 59/108 (54%), Gaps = 7/108 (6%)

Query: 35 TESDLEKNLANGMSFFV-----AEVDKIKAGVLDGFIYFFAGKHVATPGILIAEYQS 89
 T +L L+ + E+ A +D+K+ G+L+ G+ A +L+L+ Y+G
 10 Sbjet: 43 TPQELSDFLSRSTSTSIDFCLLARLDKRVGLNLNSGRV -LSQQQARADYPMIAVARTYG 101
 Query: 90 QGLGKALLKALLTEAKAQGYIK-IAMHVGNNSRAISLYQKYGFTEEA 136
 Q+G+ LL+ L A+ YI+ + + V N+AI LY+KYG E+
 15 Sbjet: 102 YGIGQLLELALDARENPYIESLKLDVQVRNTKATLYKKYGFRIES 149

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 540

A DNA sequence (GBSx0579) was identified in *S.galactiae* <SEQ ID 1725> which encodes the amino acid sequence <SEQ ID 1726>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2056 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAE14712 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 248/417 (59%), Positives = 314/417 (74%), Gaps = 4/417 (0%)

Query: 5 LALRMFRPNINEVIGQHLVGVNGKIIDRWAAANLSSMILYGGPGIGKTSIASAIAAGTTK 64
 LA RMR I ++IGQHILV KII RMW A LSSMILYGGPGIGKTSIA+AIAG+T
 35 Sbjet: 4 LAYRMRPTKIEDIIGQHLVAEDKIIIGWVAQKHLSSMILYGGPGIGKTSIALAIGATS 63
 Query: 65 YAPRTFNATVDSKRLQEIAREAKFSGGLVLLDIEIHRLEKTKQDFLLFLENGNLIIG 124
 AFR NA +++KK ++ +A+EAK SG ++L+LDE+ERLQK QDFLLF LENG II+IG
 40 Sbjet: 64 IAPKLANAVINNKIMEIVAQRAKMSGVQVILLDEVERIAKKGQDFLLFYLENGNIIIG 123
 Query: 125 ATTENFFSVTPAIRSRVQIFELFLENEDIKKAIQLAISDKERGF-PFLVTIDDEALDF 183
 ATT NP+ ++ PAIRSR QIFELFPL+ E IKA++ A+ D+ RG + V+IDD+A++
 40 Sbjet: 124 ATTANPTNAINPAIRSTQIFELFPLTPELIKQALERALEDEHKGILGTVSVSDQAMEH 183
 Query: 184 IVTATNGDLRSAYNSLDLAVMSTSPNEDGSPHISLETWNSLQCSYITMDONGDGHYDIL 243
 GD+RGA N+L+LAV+ST+ DG HI+LET E LQ + DK+GD HYD+L
 45 Sbjet: 184 FHNGCGSDVRSAINALELAVLSTKESADGSHITHTLSTASECTLQKSPSHDKIGDAHYDVL 243
 Query: 244 SALQKSRGSDVNASLHYAARLVWAGDLPSLARLTIAYEDIGLANPRAQHIVTPALEA 303
 SA QKSRGSD N+L+HY ARL+RAGDL S+ARLL +IAYEDIGL+P+A+ + A++
 50 Sbjet: 244 SAFQKSRGSDANAALHYLARLIEAGDLSEIARRLLVIAYEDIGLSPQAGPVRVLAIQT 303
 Query: 304 AQRIGFPEARILIANIVVDLALSPKNSAYLAMADALDRSSGNLPIPHLRDGHYSGS 363
 A+R+GFPEARI +AN V+L LEPKNSA LA+D ALAD+R +P+H+D HY G+
 55 Sbjet: 304 AERVGFPEARIPLANAVIELCLSPKNSAILAIDEALADIRACKIGIDVPHKLDARHYKGA 363
 Query: 364 KTLNARDYDIKYPHPEKRWVQQLYLDPLKGVNYPANETQKYKRALSGNKRIDKL 420
 + LG DYKYPH Y W+VQQLYLD L Y++ +TK+ E AL K+ DKL
 60 Sbjet: 364 QELGRGIDYKYPHINDNGWVQQLYLDPLKGVNYPANETQKYKRALSGNKRIDKL 420

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1727> which encodes the amino acid sequence <SEQ ID 1728>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2374 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 394/422 (93%), Positives = 409/422 (96%)

Query: 1 MADNLALRHRPRNINSEVIGQORHLVGNKGIIDRMVANMLSSMILYGPPGIGKTSIASAIA 60
M D:LALRMRP+ I+EVIGQ+HLVG GKII RMV AN LSSMILYGPPGIGKTSIASAIA
Sbjct: 1 MPDHIALRHRPKTTISEVIGQXHLVGBGKIIIRRMVZANRLSSMILYGPPGIGKTSIASAIA 60

Query: 61 GTTKYAFRTFNATVDSKKRLQRTAEAKFSGGVLILLDEIHRLOKTKQDFLLPLENGNI 120
GTT+YAFRTFNAT+DSKKRLQRTAEAKFSGGVLILLDEIHRLOKTKQDFLLPLENG I
Sbjct: 61 GTTRYAFRTFNATIDSKKRLQRTAEAKFSGGVLILLDEIHRLOKTKQDFLLPLENGTI 120

Query: 121 IMIGATTENPFPSVTFAIRSRVQIFLEPLSNEDIKKAIQALSDKERGFPLVTIDDEA 180
IMIGATTENPFPSVTFAIRSRVQIFLEPLSNEDIK AIQALSDKERGFPLVTIDDEA
Sbjct: 121 IMIGATTENPFPSVTFAIRSRVQIFLEPLSNEDIKTAIQALSDKERGFPLVTIDDEA 180

Query: 181 LDFIVITATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMNSLQCSYITMDONGDGHY 240
LDFIVITATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMNSLQ SYITMDONGDGHY
Sbjct: 181 LDFIVITATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMNSLQCSYITMDONGDGHY 240

Query: 241 DILSALQKSIRGSDVNASLHYAARLVEAGDLPGLARRLTITAYEDIGLANPDAQHTVTA 300
D:LALQKSIRGSDVNASLHYAARLVEAGDLPGLARRLTITAYEDIGLANP+AQ+HTVTA
Sbjct: 241 DVLALQKSIRGSDVNASLHYAARLVEAGDLPGLARRLTITAYEDIGLANPDAQHTVTA 300

Query: 301 LEAAQRIGFPEARILIANVVDLALSPKSNAYSAYLANDAALADLRSGNLPIPHRLRDGHY 360
L+AAQRIGFPEARI IAN+V+DLALSPKSNAYSAYLANDAALADLR SGNLPIPHRLRDGHY
Sbjct: 301 LDAAQRIGFPEARIPANVVDLALSPKSNAYSAYLANDAALADLRSGNLPIPHRLRDGHY 360

Query: 361 SGSKDLGNADYLYKYPHAYPEKWKVQYLPDKLVGHYFEANETGKYERALGSKNERIDKL 420
+GSK LGNA+DY YPHAYPEKWKVQYLPDKLVGH+YFEANETGKYERALGSKNERIDKL
Sbjct: 361 AGSKDLGNADYLYKYPHAYPEKWKVQYLPDKLVGHYFEANETGKYERALGSKNERIDKL 420

Query: 421 SD 422
SD
Sbjct: 421 SD 422
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 541

A DNA sequence (GBSx0580) was identified in *S.agalactiae* <SEQ ID 1729> which encodes the amino acid sequence <SEQ ID 1730>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2991 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10207> which encodes amino acid sequence <SEQ ID 10208> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 542

A DNA sequence (GBSx0581) was identified in *S.agalactiae* <SEQ ID 1731> which encodes the amino acid sequence <SEQ ID 1732>. Analysis of this protein sequence reveals the following:

10 Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2402(Affirmative) < succ>
 15 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 543

A DNA sequence (GBSx0582) was identified in *S.agalactiae* <SEQ ID 1733> which encodes the amino acid sequence <SEQ ID 1734>. Analysis of this protein sequence reveals the following:

25 Possible site: 49
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.40 Transmembrane 231 - 247 (225 - 250)
 INTEGRAL Likelihood = -9.92 Transmembrane 159 - 175 (151 - 179)
 INTEGRAL Likelihood = -9.08 Transmembrane 21 - 37 (18 - 43)
 30 INTEGRAL Likelihood = -9.08 Transmembrane 181 - 197 (176 - 201)
 INTEGRAL Likelihood = -3.35 Transmembrane 111 - 127 (110 - 130)
 INTEGRAL Likelihood = -2.81 Transmembrane 74 - 90 (74 - 93)
 ----- Final Results -----
 35 bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:CB15891 GB:299123 yxL6 [Bacillus subtilis]
 Identities = 54/203 (26%), Positives = 100/203 (48%), Gaps = 7/203 (3%)
 Query: 1 MTGLIPMKKEWLENSRSHKALALLISITIGILSPITALIMPRIMA--GILPKKIQEAI 58
 M ++ +L+KRWLE +S K + L + +I G+ PLT MPEIA G LP ++ +
 45 Sbjct: 1 MKVMALLQKRWLEGNKSGKLIWLPIANMIVGLTQPLTIYYMPRIAHGGNLPDGMKISF 60
 Query: 59 PDFTYLDYSQYQFNINQLGLILVFLFSQSIZTQREFRGTILNLTGELSKKAILIAKFT 118
 P+ + N LG+ L++F GS+ E +G ++++ ++ I++K++
 50 Sbjct: 61 TMSPSGEVWVSTLSQFNTLGMALVIPSVMGSVANERKQVITALIMSRPVTAHYIVSKYL 120

-638-

Query: 119 MFLINISYILGSLTQYATLYYFNHNGKLIIV-YGTWIFPELLLSLILFYSVIFRK 177
 + ++I +8+ G Y Y F + + G ++ + ++ L S IFR
 Sbjct: 121 IQSGVIGIMSFAAGXGLAYYYVRLFPEDASFSPASISGLYALWVIFIVTAGLAGSTIFR- 179

5 Query: 178 TAGVLJAC---LMTIVAFPISGF 197
 + G AC L V+F + F
 Sbjct: 180 SVGAAAACGIGLTAVSFAVHYF 202

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 544

A DNA sequence (GBSx0583) was identified in *S.agalactiae* <SEQ ID 1735> which encodes the amino acid sequence <SEQ ID 1736>. This protein is predicted to be ABC transporter, ATP-binding protein.

- 15 Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1344 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 25 >GP:CAB15892 GB:Z99123 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 116/303 (38%), Positives = 175/303 (57%), Gaps = 18/303 (5%)
- 30 Query: 4 ISLQNLKSPGQDIIINQVSLLEENKIYGFVGPNGAGKITTIKMILGLLKVDSGTISVM 63
 +S++L KS+ + VS + EN+ +GPNAGKITT++M+ GLL SGTI ++
 Sbjct: 2 LSIESLCKSYRHEAVIONVSPHNENECVALLGPNAGKITTILQMLAGLLSPTSGTIKLL 61
- Query: 64 GNPVTFQQTNSQVIGYLPDVPEFYDYMTAQEYLQLC---AGLAQNTSLPIADLLSQVG 120
 G + ++IGYLP P FY +MTA E+L +GL++ K I ++LE VG
 35 Sbjct: 62 GE-----KKLDRLILGYLPQYFAPYSMTANEFLTFAGLSLGSKRKCQKIGEMLEFVG 116
- Query: 121 LADN-QQRISTYSRGMKRLGLAQLIHNPKILICDEPTSLDPOGRQELLSIISQLRQ 179
 L + +RI YS GMKRLGLAQL+H PK LI DEP SALDP GR E+L ++ +L+
 40 Sbjct: 117 LHEAAHRIQGYSGMKRLGLAQLLHKPKFLILDEFPVSLDPTGRFEVLMMRELKGH 176
- Query: 180 KTVIFSTHLLSDVEKVCQDVLILTKSGIH---NLEDLRDKASASVNOALLIKVSDNEAQ 236
 V+STH+L D E+VCDQV+I+ I L++L+ + +V L++ K+ +
 Sbjct: 177 MAVLFSTHVLHDABQVCDQVIMKNGSISWKGLOHAKCQQQINVFILSVKKLKGNIIE 236
- 45 Query: 237 KIALRFPINOKDOYYKHLEISSEANNREQALAFYRYLVQSEITPYFIELLEDSLDFYL 296
 K + + + + EL + + L+ + + +T E +SLAD YL
 Sbjct: 237 KPYVSAIVYKNES--QAVFELPDIHAGRSLLSD---CIRKGLTVIRFQKTESLEDVYL 290
- 50 Query: 297 EVI 299
 +V+
 Sbjct: 291 KVV 293

There is also homology to SEQ ID 686.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.
- 55

Example 545

A DNA sequence (GBSx0584) was identified in *S.agalactiae* <SEQ ID 1737> which encodes the amino acid sequence <SEQ ID 1738>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4383 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]
 Identities = 25/60 (41%), Positives = 41/60 (67%)

Query: 2 IGDITLWERTLGMTQEKLSDYHLTKATISKWRNNQAKPDIDYLIIMAKLFDMTLDLVL 61
 +G I +R L ++QE +++ L +++ ISHVE NQ++P +D LI +A+LFD + ELV
 Sbjct: 4 LGSNISNKRKSLKLSQYVABQLGVSRQAISKWETNQSEPSMDNLIRIAELFDSIDIKELV 63

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 546

A DNA sequence (GBSx0585) was identified in *S.agalactiae* <SEQ ID 1741> which encodes the amino acid sequence <SEQ ID 1742>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4241 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAN15470 GB:Z99121 yvdC [Bacillus subtilis]
 Identities = 59/104 (56%), Positives = 76/104 (72%)

Query: 1 MDITAYQKNVSEFYKKRWVQVNSFIRSNFLCEVGEIQAIRKYEIGRDRPDEIEKSN 60
 M + +KW+ RFY+KR W +Y FIR FL IE GELA+A+R YEIGRDRPDE E S
 Sbjct: 1 MQLADAEKNMKEFYKRGWYFYGPFFIRVGFLMEFAGELARAVRAYEIGRDRPDEKESGRA 60

Query: 61 ENLNDIKKEIEGDVNLNIFLADQVNSIEIIRAHONKLEKRFE 104
 E ++ EE+GDV+ NI ILAD Y +SLE+++AH+ KL KRFE
 Sbjct: 61 EQKQELIEEMGDVIGNIAILADMYGVSLIEDVMKAHQKELTKRFE 104

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 547

A DNA sequence (GBSx0586) was identified in *S.agalactiae* <SEQ ID 1743> which encodes the amino acid sequence <SEQ ID 1744>. Analysis of this protein sequence reveals the following:

-640-

Possible site: 61
>>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0453(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:BA06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 87/187 (46%), Positives = 125/187 (66%)

Query: 1 NKITVPCGASNGRNPTYSQKIVELGEMMIKNNHDLVYGGKVGMLGVIAITVINNGQAI 60
 KCI VPCG+SNG + +Y + +LG+ + + LVYGG VG+MG +AD+V+ GG+ I
15 Sbjct: 1 NKIAVPCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGMGAVADSVLEAGSEVI 60

Query: 61 GVIPFLKDRBIAHTNLSKLIVVENMPQRKQKMSLGEAYIALPGGPGTLEEISEVISM 120
 GV+P FL++ EI+H +L+KLIVVE M +RK KM L + +ALPGGGPGTLEE E+ +H+
20 Sbjct: 61 GVMPFLKEEPETISHPHLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFEIPTVA 120

Query: 121 RIGQNDSPCLLYNINGYFNHLESMPDHVSBGFSGQDRNVLFSDDIIEIKFIKDYQS 180
 +IG + PC L NIN YF+ L ++ HM +E FL + R+ L D I + Y+
25 Sbjct: 121 QIGLHKPCQGLNLNINHYFDPPLVTLIHHMSNEQFLHEKYSMAVHTDPIILLDQSFSTYEP 180

30 Query: 181 FTIRKYS 187
 PT++ YS
 Sbjct: 181 FTVKAYS 187

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 548

A DNA sequence (GBSx0587) was identified in *S.agalactiae* <SEQ ID 1745> which encodes the amino acid sequence <SEQ ID 1746>. Analysis of this protein sequence reveals the following:

35 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5288(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 549

50 A DNA sequence (GBSx0588) was identified in *S.agalactiae* <SEQ ID 1747> which encodes the amino acid sequence <SEQ ID 1748>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence

-641-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3685 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF12706 GB:AP066865 integrase [bacteriophage TPW22]
 Identities = 106/377 (28%), Positives = 199/377 (52%), Gaps = 31/377 (8%)

10 Query: 4 ARYRRRQNLNLAAYBIRESGKTVAYNS---GPKTKKLARARAEPLQKIRGSIITKNI 59
 A +R+G W + + + Y G+TKK ARA A+ ++L S +I
 Sbjct: 2 ANFRKRGKCT--WQFRLSYKDNNGEYKCFKGGYKTKKRAAADAACKRLANNISFPNDI 59

15 Query: 60 SLPELVQSWLDLKLMPNSRSDVTKKYLGRVLEKLGDKPISQIRPSEYQIRIMNYGQ 119
 SL + ++W + P + ++ T + Y ++K DKPI++I P+ YQ ++N
 Sbjct: 60 SLDYFFERGAQVYKFP-HVTRATWRTYKRTNLNLIDKYIKDKPIAEITPTPTQAVLNKMSL 118

20 Query: 120 RVSBNFLGRLNTGVQSLQMAIADKVMIEDFTQNVELFSTVKSQDADSKYLHSEKAYLDL 179
 + L + +K +++A+ +KV+ E+F + S + + + KYL+++ YL L
 Sbjct: 119 LYRQSLDKFYFQIKSANKIADVEKVISNFADFTKRSKLAARPVEEKYLADE-YLKL 177

25 Query: 180 INAVDKRPNYKSVVFPYIIYFLKTMGRYGLLIALTWEDIDFDKGIFATYRRFN-SETSQ 238
 + ++K Y + Y TGM+ EL+ LTW +DFDK R ++ S T+
 Sbjct: 178 LAIASSMETSY---FACHTAVTGMRFAGLLGLTWSHVDFOKKEISIQRTWYSITNN 234

30 Query: 239 FVPPNNTSIRIVPVWNECLILKLNKIEQNGSNKELGLQNTNMFVPHGYPGNSVPSIN 298
 F RN++S R +P+ ++ ++LK K KE +N + V + S N
 Sbjct: 235 FASTNBSSKRPKIPISNTIKLKKYK-----KEYWHENKIDRVITNL-----SN 280

35 Query: 299 GTNKVLGIGVQELNIEPIITTKGARTHYGSLMRGTYDLGIITAKLGHKDIISMLIEVYGH 358
 G NK ++ ++ + P RH++ S+L ++G DL ++K+LGH++++ ++VY H
 Sbjct: 281 GLNKKTK-VIAGRKVP----HSLRHSFASLYIKGIDLLTVSKLGHENLAVTLKVYAH 335

Query: 359 TLEEKIQEYNEIKQLW 375
 L+E QE + I+++
 Sbjct: 336 QLKEMQSEKNDVIRKIF 352

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 550

A DNA sequence (GBSx0589) was identified in *S.agalactiae* <SEQ ID 1749> which encodes the amino acid sequence <SEQ ID 1750>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2710 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 551

A DNA sequence (GBSx0590) was identified in *S.agalactiae* <SEQ ID 1751> which encodes the amino acid sequence <SEQ ID 1752>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2534 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA05248 GB:D29979 ORF3 [Bacillus stearothermophilus]
 Identities = 81/263 (30%), Positives = 135/263 (50%), Gaps = 14/263 (5%)

Query: 65 MGHVHVLGQGCRCQYEEFIEGNDNNWTSLVKRLI-DNNSNFTRLDITANDIFDESINQVQL 123
 MG+HVE+ GQGCR +E NW L RL+ + N TRLD+A D F + L
 Sbjct: 1 MGIHVMTGQGCRLFELH---TSINWYELFRLVYEYEVNITRLDVAVDDFKGYFKINTL 57

Query: 124 VEYSKKGCLCITTARHAETHEKFPVIDSGELVGETTVPGARGNQCVYNKLMQNGKLOTD 183
 + K + + A + E VI+GE +G T+ FGA + + + E+N ++ D
 Sbjct: 58 VKKLKDEVTSRFKKARHIEIVIGGETTIGHTLYFGAPSSD---IQVRFYEKNVQMGMD 114

Query: 184 IDINSWVRAELRCWQEKANLIAHQL-NMWRPLASIFYEAINGHYRFPVSPKARDKIKRRRE 242
 ID+ W R E++ +A+++A + +D+ PL I + + +F + KA DINK+R
 Sbjct: 115 IDV--WNRTIQLRDRHVVAQITADVDLPLGETVAGLLRNVIYQFTRKATDKIKRRFP 172

Query: 243 SVRWQNYINTEKRLSIVREKPTLRQSRWTDKQVSKTIKVKYMAKYSAYGIDQAEVF 302
 R+W N++ + R++ K++ + N D QVSK+ +Y E ++ + F
 Sbjct: 173 LARFWNLPLGVDQPLRIAKQMPTSIKIKYRNIDSQVSKSPFMIYYCLNE---EEKQRF 228

Query: 303 LQDLIRRGVREKPTDNDKEIEQY 325
 + D+L G K T D + I Q+
 Sbjct: 229 IDDLVARGASKLTKADLQVINQF 251

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 552

A DNA sequence (GBSx0591) was identified in *S.agalactiae* <SEQ ID 1753> which encodes the amino acid sequence <SEQ ID 1754>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2700 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 553

A DNA sequence (GBSx0592) was identified in *S. agalactiae* <SEQ ID 1755> which encodes the amino acid sequence <SEQ ID 1756>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 50
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3121(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1757> which encodes the amino acid sequence <SEQ ID 1758>. Analysis of this protein sequence reveals the following:

```

15     Possible site: 24
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
20     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25     Identities = 19/52 (36%), Positives = 33/52 (62%)

      Query: 8  FGNLRLRKRGISQVELSNQLQIGKQISIDYEKQGAFTFANLKDIAEYF 59
                F NL L ++ I Q+++ N+L I K +I+ Y K ++ PT N+ K+A++F
      Sbjct: 15  FSTINLNLMAKJNKLQIDHKLGLGPKSTITGYVKGSLPTAGNVQKLADFF 66

```

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 554

A DNA sequence (GBSx0593) was identified in *S. agalactiae* <SEQ ID 1759> which encodes the amino acid sequence <SEQ ID 1760>. Analysis of this protein sequence reveals the following:

```

35     Possible site: 54
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
40     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

45     >GP:AAJ98584 GB:L44593 ORF536; putative [Lactococcus phage HK5-T]
      Identities = 248/532 (46%), Positives = 359/532 (66%), Gaps = 16/532 (3%)

      Query: 1  MNFIEQISENNQFPPIIFVSGGITORYFENAPTWKLLKDIWLELFDEESYAK--AFELR 58
                MNFIE I +NNQFPPIIFVSGG-T-RYF+N WE+LL ++W + +E+++Y + FE
      Sbjct: 1  MNFIEIKDNNQFPPIIFVSGGVTKRYFKNGLKWEQLLELWNLVEEEKAPTYQHVFNEL 60

50     Query: 59  ERFEN-----NDFDIYTNLASLLEKSVKAFINGNIQVNNLKLTAYPELNISPPKQLVAN 113
                + +N +F+I +A +LE++++ AF + + +INL L A+ +ISPP+Q +AN
      Sbjct: 61  LKSKNLKSDKEFEINLMMAGILEKINNAFYSDLELNINLTACARTEHISPPROCIAN 120

55     Query: 114 RFSNLKIRREKIRKIKQFSQMLSKARIITNTYDNFIESCIKNTINVSVKINVGNGGLFLK 173

```

-644-

PSNL ++ ERI PS+ML KAR I+TINYDNFIEEC NVS+K+NVGN GLF+K
 Sbjct: 121 TFSNLDKKGDFDEILISFSLMLVAKRFIVTTNYDNFIEECFSKKNVSIKVNNGSLFVK 180

Query: 174 SSDYGLYKIKHTVDDASTITIKEDYKKNVTKSALINAKILNMLVSPILFLGSLTDE 233
 S+DYGLYKIKHG+V + TI IT EDY+ N +K AL+NAKILNML ESPILF+GYSLTD+
 Sbjct: 181 SNDYGLYKIKHSGVKNNTICITSEDIYKNNESKLALVNAKILNMLTSPILFVIGYSLITK 240

Query: 234 NIRKLLTDFARNSPFDISESAQKIGVVEYLPDSESIETVVSGLPDLVSYYSLKLTNPN 293
 NIR+LLT ++EN P++ISE+A +IGVVEY PD I+ +VS++PD+ ++Y+ + TDN+
 Sbjct: 241 NIRELLTYSYENLPTFISEAAARIGVVEYTPDKIEIQDIVNIPDLGIHTTKISTDNKK 300

Query: 294 IYRLISKINGFLPSIAKIYKNVFKLIVKESKDLTVLTSYEDLANLDEIRKMI 353
 IY IS+I QG+LPSIAK+E FRKLIIVKRG+ K+L TVLTS+ D++ + +B++KMI
 Sbjct: 301 IYDELSQIBQGYLPSIAKFEAGFRKLIIVKRGKELOTLVTSFDISKINTEELKMI 360

Query: 354 VVAFGDERIYKFPDPKEVRSYFLDKETIPQEIIVIRFIATQVASHPIKKNMFAMSEY 413
 VVAFGD +YIK P +K+Y+R YF + + I +F+ + +P KK+M + +
 Sbjct: 361 VVAFGDSKIYIKMPTKDYIREYFNSMMLDTRIALLLFKRSAMTPVPYKHMVGIESW 420

Query: 414 --ISKSNKNTENIKRKLKESSELDDPTSSIGVPL--HKTLEKQTEIVGILE-ADV 468
 I D + E+K R+S E + ++ L + L + I + + ++V
 Sbjct: 421 GSIPNDLVQVESLKTRISNPFSSIVRTYSIKANKDLAKYLYLIMKTSTIEDVMSLSNV 480

Query: 469 PNVRYNFIATHIKHFPKEELFLVLEKIID----SGIFETSRRRLFAKFDLL 516
 P + PI I P EEL + K ID +GI T R+ + ++ ++
 Sbjct: 481 PLYNKLRFILFKIDKFKVZELKDFIVINIDMGEKGISSTLYRKIVMSYSII 532

A related GBS gene <SEQ ID 8599> and protein <SEQ ID 8600> were also identified. Analysis of this protein sequence reveals the following:

30 Lipop: Possible site: -1 Cread: 8
 McG: Discrim Score: 1.55
 GVH: Signal Score (-7.5): 0.27
 Possible site: 54
 >>> Seems to have a cleavable N-term signal seq.
 35 ALON program count: 0 value: 2.44 threshold: 0.0
 PERIPHERAL Likelihood = 2.44 214
 modified ALON score: -0.99

40 *** Reasoning Step: 3

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

47.3/69.1% over 531aa Lactococcus lactis

EGAD[36707] hypothetical protein Insert characterized
 50 GP[928833]gb|AAA98584.1||I44593 ORF536; putative [Lactococcus lactis phage BK5-T] Insert characterized
 PIR|T13261|T13261 hypothetical protein 536 - phage BK5-T Insert characterized

55 ORFUC184[301 - 1848 of 2154)
 EGAD[36707]38110(1 - 532 of 536) hypothetical protein [Lactococcus lactis]GP[928833]gb|AAA98584.1||I44593 ORF536; putative [Lactococcus lactis phage BK5-T]PIR|T13261|T13261 hypothetical protein 536 - Lactococcus lactis phage BK5-T
 %Match = 32.3
 %Identity = 47.2 %Similarity = 69.0
 60 Matches = 247 Mismatches = 155 Conservative Sub.s = 114

126 156 186 216 246 276 306 336
 RMLILKAFYLAKFLKYC+KK+CQTKRGOLYFRVYGLI IKIKRMVSKML**D*QLNKLIINKR+GQELVNFPIQISENNQ
 :||| | :|||
 MNFIENFIENNQ

-646-

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2933 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA98585 GB:L44593 integrase [Lactococcus phage HK5-T]
 Identities = 124/382 (32%), Positives = 202/382 (52%), Gaps = 21/382 (5%)
 Query: 1 MATYQRGRKKLWDYRIFNEKSLVA-SGSQPKTKREANNEAMRIE---QQKLLVNSISS 56
 MATY++RGK W Y I K L + GF TK +A EAM IE ++ +V+ I
 Sbjct: 1 MATYQKRGKT--MQYSISRTKQGLPRLTGKGFSTKSDAQEAMDISLAKKGFIVDPIKQ 58
 Query: 57 DITLYDL-WFENWYSLIIKPSNLAETITNKYPTFSGVIRKLFQGNQKVKIKHSAYQRKLT 115
 +I+ Y W E Y K + + E T Y ++ N ++I S+YQR LN
 Sbjct: 59 EISEYFKDWMEGY----KQNAIDEMTYKGYEQTLKYLKTYMPNLVISBITASSYQALNK 114
 Query: 116 YAEKTYKNHVRRLNSDIKKAIQFAKRDGVLLSDFDTGVVIAGRKFKVDADDKYLHSIFD- 174
 +AE + K + ++ ++ +IQ +G L DFT V+ G K DK+++ FD
 Sbjct: 115 FAETHAKASTGKFHTRVRSIQPLIEBGRQKDPFTRAVVKGNGNDKAEQDKFVN--FDE 172
 Query: 175 YKGVSYLSNLD--YENSIVYLLLL/LFKRGLRVGRLAL/TWDDWFEDEIKTYR--R 230
 YK+++ Y N L+ YS+ + + + + TG+R EA L WDD++F + IK R
 Sbjct: 173 YKGLVDYFRNRKLNPNYSSTPLFIISI--TGWASEAPGLVDDIDFNNTIKCRRTVN 229
 Query: 231 FSDGKTFSPPKTKTSITPTISQSLALILRLDQDQQLMLNKLKVINNNQIFDYDRYG 290
 + G F PKT IR I I +L+D ++ Q+ + ++ I +++ + Y
 Sbjct: 230 YRNVGSGFKPKPTDAGISRDIVDDBSMQLLKDFRQKQTLFESLGKIPHDVFCVHPYRK 289
 Query: 291 VSTNSAINSLKVLKILNINSKMTATGARHTYGSYLLAKGVDIWVARLMGHKIDITQLL 350
 + T SA+ +L + LK LNI++ +T G RHT+ S LL GVDI V++ +GH +
 Sbjct: 290 IITLSALQNTLOHALKKNLSTPLTINGLRHTASVLLYHGVDMTVSKRLGHASVAITQ 349
 Query: 351 EFNHGVLTVDNREYETVRSLV 372
 +TY H++ E+ NK+ + L+
 Sbjct: 350 QTYIHITKLENKDKDKIIELL 371

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 556

A DNA sequence (GBSx0595) was identified in *S.galactiae* <SEQ ID 1763> which encodes the amino acid sequence <SEQ ID 1764>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1603 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10209> which encodes amino acid sequence <SEQ ID 10210> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA07266 GB:AP001519 unknown conserved protein in others

-647-

[*Bacillus halodurans*]
 Identities = 26/71 (36%), Positives = 39/71 (54%), Gaps = 6/71 (8%)

Query: 37 WWDIDNLQELLGIGRSKLINDILLNPDIKKEVDLSINPNGFIYVPRKGSRYKILATK-- 94
 WW + +L+E G L +ILL+P K +D I GF+ YP+ KG R+ +A+
 5 Sbjct: 4 WWSMDLKERITGYSEDWLEENILLAPRYKPMLD--IENGGFVYYPKKGKRWCFIASSME 61

Query: 95 --ARKYFSDFM 103
 +KFP+D F
 10 Sbjct: 62 EFLKKYFDIT 72

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 557

A DNA sequence (GBSx0596) was identified in *S.agalactiae* <SEQ ID 1765> which encodes the amino acid sequence <SEQ ID 1766>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have an uncleavable N-term signal seq
 20 INTEGRAL Likelihood = -3.88 Transmembrane 12 - 28 (11 - 29)

----- Final Results -----
 bacterial membrane --- Certainty=0.2550 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB99663 GB:U57604 chromosome segretation protein (smc1)
 [Methanococcus jannaschii]
 30 Identities = 53/210 (25%), Positives = 95/210 (45%), Gaps = 33/210 (15%)

Query: 20 IFTNVLISNSRDNKAIGRELELLEBGGQELVDEFSKISTNQYDKV-----LI 69
 +F +G+L N + + + + + K++DE S I+ K LI
 35 Sbjct: 133 LFRRLGLGENVISQGLLKIINISPTERRKIIDEISGIAEFDEKKQABBEKKAKREL 192

Query: 70 Q-----SNLSNNVIERKNQELVQKNSYVK--EDTKYIRDENLIRKSK-----BEVYNHV 116
 + S + NN++K K+E Y+K E+ K + +++K S E + N +
 35 Sbjct: 193 EMIDTRISEVENLKKLKKKEKEDAEKYIKLNEELKAAKTALILKVSYLNVLENIQNDI 252

Query: 117 KNGDKLIEKMAFANELILKFGVSRNQMLGLKVNLSLEKIVDLNCPKNDIEIKRKEI 176
 KN ++L NE + K E+ E + L L++N+ I++ N+ N+E+ +L ESI
 40 Sbjct: 253 KNLREL-----KNELPLSKVREIDVIEINLKLRLNN----IINELNKKGNBEVLEIKSI 302

Query: 177 SSFERELSRFQVGYSEAEISKSTLARRILN 206
 E E+ + V S E+K I N
 45 Sbjct: 303 KEILVREINDKKVLDSINELKKVVEIEN 332

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1766 (GBS315) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 4; MW 26.7kDa) and in Figure 239 (lane 5; MW 41kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 5; MW 52kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 558

A DNA sequence (GBSx0597) was identified in *S.agalactiae* <SEQ ID 1767> which encodes the amino acid sequence <SEQ ID 1768>. This protein is predicted to be surface protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 26
   >>> Seems to have a cleavable N-term signal seq.
       INTEGRAL    Likelihood = -7.70    Transmembrane  229 - 245 ( 226 - 248)

10  ----- Final Results -----
       bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:QJ47097 GB:X66468 orf iota [Streptococcus pyogenes]
       Identities = 90/262 (34%), Positives = 138/262 (52%), Gaps = 26/262 (9%)

       Query: 4   VKVLSLITV-SGLFIMAGNLSASADVVISGGCTIMLSGVDAGVSCSIMPFPSSINPV--- 59
               +K L+L+T+ S L++ ++ AD S D +L+ D V P ++PV
20  Sb|ct: 1   MKKLALLTLPTLLVSAPIVSFADETASSSDINILADDPVVPVEPTDPTTVPDVPDV 60

       Query: 60  -----TDTTESAPTPTSDPI--TDTTESAPTPTSDPI--TDTTESAPTPT 104
               T+ TEP+ PT T+P T+ TEP+ PT T+P T+ TEP+ PT T
25  Sb|ct: 61  DPVDVFPVDPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 120

       Query: 105 DQTITGTTDSS-TPSSSTTNFVGITDNGTKPNAGIDKFPSTNKPSTHSESSI--KPVTKPT 161
               + T T + T S T P + T+P + +PS +E ++ KPV
30  Sb|ct: 121  EPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 178

       Query: 162 INQPTITTVTGDQVIGTQDGKVLVOTPSGTQLK-DAAEVGNVQKRDGTVAIKKSDGKIEVL 220
               I P+ T TG ++ +D K ++Q GT K +A E+G +VQKDGTV +K SDGK++VL
35  Sb|ct: 179  IENPVNTDGTGVVIVAVEDSKPIIQLADSTTKKVEAKEIGADVQKRDGTVTVKSGDGMKVL 238

       Query: 221 PKTGGKRTI-FTIVGLLLIAGA 241
               EKTGE I +++G L++ G+
       Sb|ct: 239 PKTGETANTALSVLGSLMWLGS 260

```

There is also homology to SEQ ID 760.

SEQ ID 1768 (GBS141) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 4; MW 35kDa). The GBS141-His fusion product was purified (Figure 19, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 295), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 559

A DNA sequence (GBSx0598) was identified in *S.agalactiae* <SEQ ID 1769> which encodes the amino acid sequence <SEQ ID 1770>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 18
   >>> Seems to have a cleavable N-term signal seq.

       ----- Final Results -----
       bacterial outside --- Certainty=0.3000(Affirmative) < succ>
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8601> and protein <SEQ ID 8602> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 4
McG: Discrim Score: 14.39
GVK: Signal Score (-7.5): -1.23
    Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 8.96 threshold: 0.0
    PERIPHERAL Likelihood = 8.96 104
    modified ALOM score: -2.29

*** Reasoning Step: 3

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 1770 (GBS17) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 2; MW 24kDa).

The His-fusion protein was purified as shown in Figure 189, lane 10.

Example 560

A DNA sequence (GBSx0599) was identified in *S.agalactiae* <SEQ ID 1771> which encodes the amino acid sequence <SEQ ID 1772>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS gene <SEQ ID 10779> and protein <SEQ ID 10780> were also identified. A further related GBS nucleic acid sequence <SEQ ID 10957> which encodes amino acid sequence <SEQ ID 10958> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1772 (GBS643) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 2-4; MW 79kDa) and in Figure 186 (lane 2; MW 79kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 5-7; MW 54kDa) and in Figure 176 (lane 5; MW 54kDa).

GBS643-GST was purified as shown in Figure 236, lane 7.

-650-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 561

A DNA sequence (GBSx0600) was identified in *S.agalactiae* <SEQ ID 1773> which encodes the amino acid sequence <SEQ ID 1774>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5815 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 562

A DNA sequence (GBSx0601) was identified in *S.agalactiae* <SEQ ID 1775> which encodes the amino acid sequence <SEQ ID 1776>. This protein is predicted to be membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -13.32 Transmembrane 311 - 327 ( 282 - 332)
INTEGRAL Likelihood = -10.46 Transmembrane 293 - 309 ( 282 - 310)
INTEGRAL Likelihood = -8.55 Transmembrane 390 - 406 ( 388 - 410)
INTEGRAL Likelihood = -7.64 Transmembrane 49 - 65 ( 40 - 69)
INTEGRAL Likelihood = -5.68 Transmembrane 100 - 116 ( 98 - 122)
INTEGRAL Likelihood = -4.35 Transmembrane 130 - 146 ( 127 - 148)
INTEGRAL Likelihood = -3.88 Transmembrane 344 - 360 ( 342 - 363)

----- Final Results -----
bacterial membrane --- Certainty=0.6328 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB70618 GB:A243106 membrane protein [Streptococcus thermophilus]
Identities = 234/665 (35%), Positives = 379/665 (56%), gaps = 59/665 (8%)

Query: 13 FAKVKDVIDFALKAYMEITH-GARTGACSIILLDFVNFPPFFLLINLIVGLFVILRFFENF 71
FAK+K VDIF+LK+YME T+ G+ GA ++ ++FVN FF+LN +VG PS++R E
Sbjct: 5 FAKLKGVDIFLSKSYMEPTNFGSFGNAWVLINLFFVNLFFFLINAVVGFFSLIRILEKI 64

Query: 72 SLVITYKQIVYHSSQKLWNLGN--GSYTS-SLLYLVAISAPISIPISYLFSEKGFPSKR 128
LY TYK V+H + +W +G+ G T+ SL+ L+ + AF +F Y FSKG PS+
Sbjct: 65 DLYATYKTYTVFPGASSIWNHGFSGTNGTNTKSLNGLWTLNLLVAFILFYQYFSEKGSFST 124

Query: 129 LIHLFVWIIIGMGYFGFTIQSTSGGIYILDTHVHNGSPSDAVINLSLNPSSGKTKITOK 188
L+H+ V+L+L +GYFOT+ TSGG+Y+LDTV+ ++ + + +D KI +
Sbjct: 125 LIHLFVWIIIGMGYFGFTVAGTSGLLYLDTVHNVSKDVTKIAGIKVDYAKIKSIKIGK- 183

Query: 189 SSVADNYVMKTSITYATLFPVNTQGLNGKPFHNNQTKREKFDNQVQLKYDKSGKPFITPEOK 248
S++D+Y+ +TSY AY+FVNTQQ NGK+ N+Q GKRE FD+ +VLG DK+G F K K
```


-651-

Sbjct: 184 -SMSDSYIAETSYKAYVFNVTGQENSGKYKNSQDGKEAFDGSKVLGTSKNGNFGAKAK 242
 Query: 249 DLIANYDNIQKATREGEKKRWLSAVNDYLMIESGVILKIFPAVILAVPLILICLIAPM 308
 + Y D+LG+ A+ EKRW+SA+ D+++ XVI KI EA +LAVP+ILIQL+ +
 5 Sbjct: 243 HRSKYLDDLGEGANDGEKRWVSAMPOFIPTKVFVIVIPKIVRAVILAVPILICLIAPM 302
 Query: 309 ADVLVILIMFIFPLALLVSFLPRMQDILIPWLVKVMFGAVSPALAGFLITLIVFYQTLIA 368
 A +LV+ ++ +FP+ LL+SP+PRMQ+++F VLVKVMFG + FPA+ LTL++FY + +I
 10 Sbjct: 303 AQLVLVIMLILFVIVLMSFVPRMQELVFGVLKVMFGGLIFPAITLILITLIFYTEKME 362
 Query: 369 TFFVKKFTDGSLLGSNFKGQAILFMILLITVFFVQSVFWSIMVYKKTFLRLILIGSRASQV 428
 V F DG L + + ++F LL++V +G +++ IW++K L+ I+GS+A V
 15 Sbjct: 363 NIVINGF-DGVLKTLPSLLPLGLVFKLLVSVSGVYIPLIWRFGQQLLQFILGSKARMV 421
 Query: 429 -----INQSVQKINEKARNLGITPKSIYERANDMSSLAMGAGYGVCTNQAQ--DN 478
 + V K E A + P A + + + GAG+G G MNA+ N
 Sbjct: 422 ATDGTGKEHGVTKSEVASQV---PTRSLATAGELGNPTLAGAGFTGVMSAKSHFQN 478
 Query: 479 WIAFKERQANLDDQSKNTADKYDEANADDVTISKESALINBEGEYQSILPKESKRIE 538
 +F R++ + + + + + + +I
 20 Sbjct: 479 AGSFPTTRKPSQPTVMPGPTAPITPESPEPIIP-----PTCTPDNFKTIG 527
 Query: 539 QLGKESYSELSPISGNSREILKVNKSDNHTPQSGDGTSLTNQOMITNDIENHSNNT 598
 + + +SEG + E + + +
 25 Sbjct: 528 EEKFTPDSPIMSDPTPSE-----DEPQTLKEEM 559
 Query: 599 SPLQKRLNKLGEGLSQPNDSVNTQNHGKNAFEKGHASKTKEVKQHNLERQSKVLEE 658
 SP KQ ++N LE L + +M K G NAF + + T+ + + N+R+ ++ +
 30 Sbjct: 560 SPFQHRINTLERLDAYKDQAMVYKQGSNAFTAYRKTLTRDKIRANIRERRRLTQR 619
 Query: 659 LEKLR 663
 L +R
 Sbjct: 620 LNLQR 624

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 563

A DNA sequence (GBSx0602) was identified in *S.agalactiae* <SEQ ID 1777> which encodes the amino acid sequence <SEQ ID 1778>. This protein is predicted to be conjugative protein. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3714 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CB70617 GB:AJ243106 conjugative protein [Streptococcus thermophilus]
 Identities = 515/757 (68%), Positives = 612/757 (80%), Gaps = 1/757 (0%)

Query: 1 MSDFEADLADDVKELGLSTLDFTVDITLHEMELPYQFDWLGVDLKGQYNANIKETYN 60
 M DF LADD +ELG E L +TVD LT EMEIPYQFDW+IGV L K + A +K+ Y
 55 Sbjct: 78 MRDPSEALADDSRELGEELLTLVDRLTDEMELPYQFDWLVGVTLKQNHGATPYKDLAYS 137
 Query: 61 QFESIASNFASLAGYEVEVDWYKHSSEELLVYSLSTLAKRLTVDVLFYQRMQFL 120
 F + A GYE + WY ++ +E ++ S L+AKRLT++LFFYQRMQ+L
 60 Sbjct: 138 SEFSEKIALKGLGYEYALSPTWYDYRSDEFTIFQAFSVLRKRLTNEELFYQRMQYL 197

Query: 121 RYVPPTKSEVIARNAFLVDTTLIKSLGGFKLESAYGSSPVSVLPVGRPSTFINGPHL 180
 RY+PH K EV+ARR+ N+TTLIK L+GGFL+LES YGSSPV+LPVG+P INGPL
 Sbjct: 198 RYIPHYKGEVLNRSQFNITLTKLVKGGFKLESAYGSSPVVILPVGKPPVQMGPHL 257

Query: 181 GELVQRMSPFPVELRKFARFIDCKLQGTMGSRNTRYDQIMKKEAYNTINTVQQDILMAYS 240
 GE VQR++FPVELR KAFID K+ G MGRSNTY IM+EA NT+VQQD+I+MG+ S
 Sbjct: 258 GEFVQRIAPFVELRIKAEFIUTWKIKGRMGRSNTYRHNIMEAENTDTVQQEIRMGIS 317

Query: 241 LKDLMKKVGKNEKIEIYGYLVVAGSSNLQKORRYALLSYFDMDKNNVYRASHOTPYLF 300
 LKDLMKKVGKNE+IIEYG YL+V+ SS+NQL+QRR IL+YFDOM V+ EAS D PYLF
 Sbjct: 318 LKDLMKKVGKNEKIEIYGYLVVSSNVNQLQRRQVILNYFDOMGVVEISKASQGPYLF 377

Query: 301 QALLYGQDLQRTTHKWHILVTARGFSELMFTNTQSGNRIGWYIGRVNRLTANDSIDEA 360
 QALLYG++LQK TR W H+VTARGFSELM FTNT SGNRIGWYIGRVDN + WDSI +A
 Sbjct: 378 QALLYGENLQRTTWTWHMVTARGFSELMFTNTSSGNRIGWYIGRVNRLTANDSIDEA 437

Query: 361 IMGSKNLVLPNATVANKEDVAGKVKNPVHITGATGQSKSYLAQMIPLHTAQCNRVLY 420
 I SKN+VL+NATV NKRD+AGK+TKNPVHITGATGQSKS+LAQ+IFL A QNV+ LY
 Sbjct: 438 IDSCKNLVLPNATVANKEDVAGKVKNPVHITGATGQSKSFLAQIIFLSVALQNVATLY 497

Query: 421 VDPKRELQHYLYKVSDEPSYAKFPLEKKQIBETNFVTLDSVKNHGVLOPIVLDKRG 480
 +DPKRELH HY +V++ PE+AR++P RKKQI+ NFVTLDS+ NHGVLOPIV+LDKRG
 Sbjct: 498 IDPKRELKNHYQEVINSPEFARYPERKKQIDNFNFVTLDSLPNKHGVLOPIVLDKRG 557

Query: 481 ASSTAKMLLYLLGNATEIKLDQTTALTALSCVIAKREAGEVVGQNVIEVLIDSSDE 540
 A AKML +LL+ ++ +DQ TA+TEAI+ ++ +R AGE VGF V+R L ++ S E
 Sbjct: 558 AVEVAKMLSLFLQAVDDVINDQKTAITAINFIVERRVAGENVGPHVLETLNASSDE 617

Query: 541 VQSVGRYFKAIQNSILELAFSDGQVAGLSYBERVTVLEADLSLPKDGSDHISDHSS 600
 + SVGRY +I+ NSILELAFSDG GL+YE RVI+LEV +L LPKD S ISDHE NS
 Sbjct: 618 IASVGRYLTSTVINSILELAFSDGTTGLNYESRVTILEVNLKPKDSTKSDHENS 677

Query: 601 IALMFALGAFCKHFGERSDCE+TVEIFDZAVLMQSSBGKAVIKSMFRVGRSKYNVLMV 659
 IALMFALGAFCK HFGER+++E T+E FDSAV+LM+S+BGKAVIK+MRR+GRSK N I L+
 Sbjct: 678 IALMFALGAFCTHFGERNENEDTIEFFDZAVLMKASBGKAVIQMRRIGRSKNTALAI 737

Query: 660 QSVHDAENDDDTTGFGTIPFYEKSEREDILSHVGLVETPNLEWINNMISGQCLYDYV 719
 +QSVHDAENDDDTTGFGTIPFYEKSEREDIL HV LEVT NLEWINNMISGQCLYDYV
 Sbjct: 738 TQSVHDAENDDDTTGFGTIFAFYEKSEREDILRHVNLEVTSSNLEWINNMISGQCLYDYV 797

Query: 720 YGNLNMISIHNIHPDIDPLKPKMTVSSHLENKYAS 756
 YGNLNMIS+HN+ DID LLKPKM TVSS LENKYAS
 Sbjct: 798 YGNLNMISVHNLFEIDIMLLKPKMTVSSSHLENKYAS 834

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 564

A DNA sequence (GBSx0604) was identified in *S.agalactiae* <SEQ ID 1779> which encodes the amino acid sequence <SEQ ID 1780>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.3469 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

-653-

>GP:CAC18595 GB:A7278419 IS1381 transposase [Streptococcus pneumoniae]
Identities = 110/125 (88%), Positives = 119/125 (95%)

5 Query: 61 MYFEASKQLTDVRPKRLGVQRTTFEEMLAVLKATYQKHAKGGRTPKLSLEDLLMATLQ 140
MYFEASKQLTD RPKRLGVQRTTFEEMLAVLKATYQ KHAKGGR PKLSLEDLLMATLQ
Sbjct: 1 MYFEASKQLTDARPKRLGVQRTTFEEMLAVLKATYQLKHAAGSRPKLSLEDLLMATLQ 60

10 Query: 141 YMYEYRTYELAADFGIHESNLIIRRSQWVESTLIQSGPTISKTHLAEFTVIDATEVKI 200
Y+MYEYRTYELAADFG+IHESNLI+IRRSQWVE TL+QSG TIS+T LS+EDTV++DATEVKI
Sbjct: 61 YMYEYRTYELAADFGVHESNLIIRRSQWVEVTLVQSGVTISRTPLSSEDVTVIDATEVKI 120

Query: 201 NRPKK 205
NRPKK
Sbjct: 121 NRPKK 125

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 565

- 20 A DNA sequence (GBSx0605) was identified in *S.agalactiae* <SEQ ID 1781> which encodes the amino acid sequence <SEQ ID 1782>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -12.58 Transmembrane 39 - 55 (32 - 66)
25 ----- Final Results -----
bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 566

A DNA sequence (GBSx0606) was identified in *S.agalactiae* <SEQ ID 1783> which encodes the amino acid sequence <SEQ ID 1784>. This protein is predicted to be Cag-W. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.82 Transmembrane 50 - 66 (49 - 66)
INTEGRAL Likelihood = -3.72 Transmembrane 25 - 41 (23 - 45)
45 ----- Final Results -----
bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 567

A DNA sequence (GBSx0607) was identified in *S.agalactiae* <SEQ ID 1785> which encodes the amino acid sequence <SEQ ID 1786>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -7.80 Transmembrane 36 - 52 ( 32 - 60)

----- Final Results -----
bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12298 GB:Z9106 similar to transposon protein [Bacillus subtilis]
Identities = 68/339 (20%), Positives = 133/339 (39%), Gaps = 49/339 (14%)

Query: 16 KKEGGKQKPKTEKVKQRTANFIV--YQILGLLFTVGFPGSLRAIGLSNQVHLKETVIAV 73
K+ E ++ K K + R+ V + + G L + L + I + Q+ + K+
Sbjct: 24 KRIRPEKDKQKVPRDRSKLIAVTLWSCVGGLLFICLLAVLSINTRSQLNDMKDETNEP 83

Query: 74 EKKEGKKKTTDDSLDISRIQYMMNFVYYYINYS--QTADQKTELENY-----YSF 123
K K ++ + + + + P+ Y+N Q++ ++R L+E Y +
Sbjct: 84 TNDCKKQ-----ISVTAENFLSGFINETMVKNDQESTIEKRMQSLRSYVMVKQBNHFED 138

Query: 124 STASKTDDVRKSRITLQTRLLSVEKEKDYIALMRIGYEV----- 163
D ++ R L+ L +V++ + ++ YE
Sbjct: 139 EERFNVGLGKDELKGYSLYNVKGDKNSLFPQYKVTYENLYPVEKEVEKVKDGKKKKK 198

Query: 164 -----DKESYQMLAVFPQMQRGLLAIVSQPYTVAEDLYLGSKAFKKITLDQVKEL 215
+K QM L +P + A+ + PY +Y K K + E
Sbjct: 199 VKEKVTNEKYKQMLNITPVTKGDSFAVSAVPYFT--QIYDLKGDIAPKKEETREDEY 256

Query: 216 SKQVSSIQKFLPVFPFKYALINKITDLKLLMKTPELMGHGFVKSELDLNNAIYQEKKHQ 275
+ E+ SI+ FL PP KYA K ++ +MK PE + E + + ++ KK
Sbjct: 257 AGEKKSIESFLQNPFKCYASEKKEEMVYMMKKPEALEGNLLFCE--VQSVKIFETIKGF 314

Query: 276 VVQLSVTFDELVTGGTRSENFTLYLFPKADNGWYVEMYH 314
V +V P++ +E P+L + + +YV ++ H
Sbjct: 315 EVFCAVRFKKENDIPVNEKPSLEITENSGQFVNNKLKH 353
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1786 (GBS333d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 8-10; MW 58kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 11 & 13; MW 33kDa), in Figure 182 (lane 2; MW 33kDa) and in Figure 185 (lane 3; MW 58kDa).

GBS333d-GST was purified as shown in Figure 236, lane 2.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-655-

Example 568

A DNA sequence (GBSx0608) was identified in *S.agalactiae* <SEQ ID 1787> which encodes the amino acid sequence <SEQ ID 1788>. Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4177(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB38326 GBS:Y17736 hypothetical protein [Streptomyces
      coelicolor A3(2)]
15      Identities = 45/80 (56%), Positives = 56/80 (69%)

Query:  4  FTERANKDYVSNQEDKKILKINRLIIDIKRDPFEGIGKPELKYHYSGAMSRRITEH 63
      FT  M+DYV M + D+K+ KRINRLI DI RDPF+G+GKPEPLK SG WSRRI + H
Sbjct:  5  FTSHGKEDYVHNASDRKVTKRINRLIADIARDPFKGVGKPEPLKGDLSGYWSRRIDITH 64

20      Query:  64  RLTYMIEDGEIYLSFRDHY 83
      RL+Y  D ++ + R HY
Sbjct:  65  RLIVYKPTDDQLVIVQARYHY 84

```

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 569

30 A DNA sequence (GBSx0609) was identified in *S.agalactiae* <SEQ ID 1789> which encodes the amino acid sequence <SEQ ID 1790>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.5669(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 A related GBS nucleic acid sequence <SEQ ID 10211> which encodes amino acid sequence <SEQ ID 10212> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AADL7306 GBS:AF121418 putative Pld protein [Francisella
      tularensis subsp. novicida]
45      Identities = 26/84 (30%), Positives = 45/84 (52%)

Query:  4  MEALVYSHFRNNIKDYMKVNDSEFPLIVVKNKPDENIVLSQDSWESLQRTIRIMENDY 63
      M+ + YS FRN L D M +V  P+IV + E +V++S + +++ +ET IM +
Sbjct:  1  MQTVNYSTFRNELSDSMERVTKNHSFMIVTRGSKKEAVVMSLEDFAVEETAYLMRSNI 60

50      Query:  64  LSHKVVINGISQVKEKQVTHGLE 87
      ++ N I +V+ + LIE
Sbjct:  61  NYKRLQNSIDEVSGLAIQKELIE 84

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 570

- 5 A DNA sequence (GBSx0610) was identified in *S.agalactiae* <SEQ ID 1791> which encodes the amino acid sequence <SEQ ID 1792>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2407 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

- 15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 571

- 20 A DNA sequence (GBSx0611) was identified in *S.agalactiae* <SEQ ID 1793> which encodes the amino acid sequence <SEQ ID 1794>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.1274 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

- 30 A related GBS nucleic acid sequence <SEQ ID 10213> which encodes amino acid sequence <SEQ ID 10214> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- ```
>GP:AA660015 GB:U09422 ORF18 [Enterococcus faecalis]
Identities = 41/140 (29%), Positives = 73/140 (51%), Gaps = 3/140 (2%)

35 Query: 23 FPEVMSRLKALGLAREDDLYITADESDCQL-LKRDHSTFMTNQFVELVIANVDSLVKAV 81
 PP++ E+K +GL +E + EY I D + + E+ SI +N+ E+V + EL +
 Sbjct: 26 FPIDFEEVKKIKGLANDEYE-KYATHDYRLPTVDYETSTIGELNRLNEVUSLPEELQSEL 84

40 Query: 82 HQVIGYFASDFVDYDFNFGDCCLSDVTTRELGEYFFDELGVQGVGKEALKNYFDHAY 141
 ++ + +S + + D + SD ++ YY +E G G +L+ Y D++AY
 Sbjct: 85 SALLTHPSS-IRELSHQEDIIHSDDCMYDVARYYIEFTGALGEVPASLQNYIDQAY 143

Query: 142 GRDIDLESQGQGFSDYGYVKI 161
 GRD+DL ++G EI
45 Sbjct: 144 GRDLDSGTFFISTNIGIFEI 163
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 572**

A DNA sequence (GBSx0612) was identified in *S.agalactiae* <SEQ ID 1795> which encodes the amino acid sequence <SEQ ID 1796>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1366 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 573**

A DNA sequence (GBSx0613) was identified in *S.agalactiae* <SEQ ID 1797> which encodes the amino acid sequence <SEQ ID 1798>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1484 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 574**

A DNA sequence (GBSx0614) was identified in *S.agalactiae* <SEQ ID 1799> which encodes the amino acid sequence <SEQ ID 1800>. This protein is predicted to be abortive phage resistance protein. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2205 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10215> which encodes amino acid sequence <SEQ ID 10216> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB53710 GB:U94520 abortive phage resistance protein
[Lactococcus lactis]

```

-658-

Identities = 131/499 (26%), Positives = 210/499 (41%), Gaps = 97/499 (19%)

Query: 3 MFSKIEFKNFMSFSNLT-----FDLNRGKCKDIIAIYGENSGKCN 44  
 M F+NF+SF L+ D+ N K + IY N SGK++  
 5 Sbjct: 1 MIVNFRFENFLSPFLSTFSPMAGKSHQHMEDLIELDKKQRLKLTITIGANASGKSS 60

Query: 45 IVEAF---KLVL-----SLQSMRLANENYRLQSLLEKQTNKE---ENQKINFDIISHL 93  
 V+A K L++ L S N+NT SL + + E ++ ++G S IL  
 10 Sbjct: 61 FVDAIGISKSLITIGFYNGVLGSLNSYNTNVDNLSNEFKFEYELVIEDKVYSTG-FSVIL 119

Query: 94 DKISFPTTFKGLAGNTHRIASEGNITLKYYFNIEKONGYYLLEYNNENNLKVELVFKIK 153  
 F + + N ++ Y KDN YN N+E L +  
 10 Sbjct: 120 SLKKPMSEWLYDITNDEK-----IYIDRNK-----STINDEP---LNLDEQ 161

Query: 154 SNKGVHFSITNIDGLSQSLNKTIFKNITPKDLTEQLEKYWGKHTFLSIFN--MYCLEV-- 209  
 SN + I + S + N +F N++ D + IE F +FN N LEV  
 15 Sbjct: 162 SNRISIIYDD---SANDNTQLFLNL-NDGKKTIESKDNSTIFKGVNWFNLTLEVLG 216

Query: 210 -----NEEP---INEQVSINPQKVDEPDKIPWISNFRGPFHSTELLK 251  
 REF + + + +N V+D N F E +L  
 20 Sbjct: 217 PGDEANGSIASLTQEBEEFKEDLGKYLELNDTCVIDIVQVPVDMNSV--PAKLERID 274

Query: 252 DISKQKIDBEKEKLSYTEELIYKYPALYIDIKVKQKQDAQQCEIKYELMIRIHIGD 311  
 +I+ I K+K+ E I F+ + ++ Q+ Q +EL K+ G  
 25 Sbjct: 275 NITT-DIKKKKKR-----EDIEISFWITLNTSNIYIIQNDQCFEYFLKP-KHNGT 327

Query: 312 LLDVFISLESQOTGLDLKLV-FRWVLGDKICTIDEISDGHDLNANSTINDLK--GSV 368  
 L +S ES GT L++L V F+N D K+ ++DEID +H L+ + K S+  
 30 Sbjct: 328 LYS--LSEESDGTVRLIELFSLVFN--DEKVPVIDEIRSLHPLLTNPYESFKKQSI 383

Query: 369 NQLIPTTHDTLL--KELSPSSAYFLNIDKNGKVIISGNEADKIGVANNLEKLYLGG 426  
 N QLI TTH+ +L + L +F++ +GN +S E ++ ++ YL+G  
 35 Sbjct: 384 N-QLIVTTHDYILNPELLRDEVNFVDKNGNSMPSLEEFKERT--DKDINTSVLGG 440

Query: 427 FPGAVPDLFDIDFSLFD 445  
 +G +P L FS+ D  
 35 Sbjct: 441 RYOGIPN-LCLSPFAED 458

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 575**

A DNA sequence (GBSx0615) was identified in *S.agalactiae* <SEQ ID 1801> which encodes the amino acid sequence <SEQ ID 1802>. This protein is predicted to be repressor (rstR-1). Analysis of this protein  
 45 sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3724 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

55 >GF:AAB84427 GB:AF027868 transcription regulator [Bacillus subtilis]  
 Identities = 31/81 (38%), Positives = 53/81 (65%), Gaps = 2/81 (2%)

Query: 9 QKLKELRKEKKLTCTELASKINISOKSYSNWSSGKAEPTLDNIKLANILDNVVDYLG 68  
 Q+L+LRK KLT +LA K+ I++ SY +E+ +P LD ++ LA + DV+VDY+LG  
 60 Sbjct: 4 QRLRLKRAKHLITMQLAKTIGIAKSSVGGYAEKSPFLDKVLILARLYDVSVDYIG 63



-659-

Query: 69 SDNFSNTIVLSKQNMKSPSKR 89  
 -D+ + + N+K F ++  
 Sbjct: 64 TDDPDPKV--RRNLKEFLEK 82

5

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 576

- 10 A DNA sequence (GBSx0616) was identified in *S.agalactiae* <SEQ ID 1803> which encodes the amino acid sequence <SEQ ID 1804>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

15

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3607(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 577

- 25 A DNA sequence (GBSx0617) was identified in *S.agalactiae* <SEQ ID 1805> which encodes the amino acid sequence <SEQ ID 1806>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

30

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0564(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 10217> which encodes amino acid sequence <SEQ ID 10218> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12294 GB:Z99106 similar to transposon protein [Bacillus subtilis]  
 Identities = 93/348 (26%), Positives = 164/348 (46%), Gaps = 28/348 (8%)

40

Query: 81 SRLQVMIDYVRITLKDVRDLFFPCRNFLHCAFKRPQPFESKIMNMYHLWGRGDIWFDFA 140  
 S L M+DY+R+ K D++ LE + +S Y ++ I +F A  
 Sbjct: 26 SPLVMVDYIRVSFK-THDVDRIIEVLELSKDFMTEKQSGFYGVGYLGYIKVPYSA 84

45

Query: 141 DKHETGNFQITVQLSGRCRQLELMETKPTWHDWLSYLRNSYRDMNVTRFDIAIDEL 200  
 G + +++SG+GCRQ E +E K TW+D + ++ + + TRFD+AI+D  
 Sbjct: 85 PDDNRG---VLIEHSGGCRQRFESFLECKKTIWYD---FFQDCMQQGGSFTRFDLAIDD- 137

50

Query: 201 YLGKRENEQPHLSDMISKYRYRHELDFFSLRTVWNYIGGSLNFSUMSEIEQNQRGISLYP 260  
 + F + ++ K + E R ++ GS + SD G ++YF  
 Sbjct: 138 -----KRTYFBIPELLKCAQKQGE-ISRFRKSDP--NGSFDLSD-----GITGGITTYF 183

Query: 261 GRSQSEHSEYKPKYKKEIAKQGIYKSEFALRILFELNNRYEYRLGSGKNAAVDFEISFGV 320  
 ++S+E Y FYEK YE A ++ E++ + WNRYYEHL +A A+D +  
 Sbjct: 184 GSKKSEAYLCFYKKNYBQAEKYNLPLEESGD----WNRYYELKNRSEQAVADALLKTKD 239

Query: 321 IGEISGLVLSIDVDYDQKNKY--GSGQKRLQWMPGGVPELFKFTYKPAETSESTLR 378  
 + I +I+D +D ++ W G V L KP+ ++ W  
 Sbjct: 240 LTLIAQIINNNRVFVADENITNRYEYKPLFWSFDIGVGRGLVLPKFDQFQSRNN 299

Query: 379 LSDSVSSMLLRMYIMIDVDYGLVTIANSGKVREREGIKSLKSIASL 426  
 L S +P++M+LEADH + L ++ E++ +x+L A +  
 Sbjct: 300 LRNSCAPTMWLEADHFKITLSDMIAEAEADKHKKMLDVMYDAD 347

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8603> and protein <SEQ ID 8604> were also identified. Analysis of this protein sequence reveals a RGD motif at residues 131-133.

The protein has homology with the following sequences in the databases:

29.4/54.5% over 342aa

EGAD108511|hypothetical protein Insert characterized OMV|NT01B5C566 conserved  
 hypothetical protein Insert characterized  
 GP1881297|doj|BAA19324.1|AB001488 SIMILAR TO ORF20 OF ENTEROCOCCUS FAECALIS TRANSPOSON  
 TN916. Insert characterized  
 GP2623787|emb|CB212294.1|Z99106 similar to transposon protein Insert characterized  
 PIR|G69774|G69774 transposon-related protein homolog ydcR - Insert characterized

CRP00101(205 - 1581 of 1887)  
 EGAD|108511|BS0487(6 - 348 of 352) hypothetical protein [Bacillus subtilis] OMNI|NT01BS0566  
 conserved hypothetical protein GP|1881297|dbj|BAAL9214.1|AB001488 SIMILAR TO ORF20 of  
 ENTEROCOCCUS FAECALIS TRANSPOSON T916. (Bacillus subtilis)  
 GP|2632787|emb|CB12294.1|J299106 similar to transposon protein [Bacillus subtilis]  
 PIR|G69774|G69774 transposon-related protein homolog ydCR - Bacillus subtilis  
 #Match = 9.7  
 #Identity = 29.3 #Similarity = 54.4  
 Matches = 103 Mismatches = 146 Conservative Sub.s = 88

153            183            213            243            273                                 489            519  
GVV\*RSENGHAGHSATRALQRLDSLILKPPPSNKGVRNEKPRILTPKNLYVSVRFR-----EQGRKLTLFKYQETIKSHFG  
                                        : :| | : |||  
                                        MDELQPPHANRGV-----  
                                        10

567            597            627            657            687            717            747  
YLVV--ENDS--SRLLQWIDYVRITLKDVRLEFFCRNFLCAKKEQFPFSSKAYNHLWRGDIWIIFADPKHSTGNFC  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
VIVVEKNVESPLVMVSVDYRVSPK-THDVRLIEEVLHLSCKDFMTRKGCGFYGVGTVELDYIKVFYSAPDNRG----

30                  40                  50                  60                  70                  80                  90

1017 1047 1077 1107 1137 1167 1197 1227  
LRTWYIGGSSISNDSMEETENRQGISLYPGSRQSEMYNFKRYETAKQSGITVERALKIFELWNRYSIRLSQSKAN  
: : : |||::| : : |||::| : : ||| : : ||| : : ||| : : |||::| : :  
FRKSDF--NGSDFLSD---GITGGTITLYPGSKSKAYLFYKQYENQAEKYNIPLE---LGDWNRYSIRLKNRQA  
170 180 190 200 210 220

1257      1287                      1341              1371              1401              1431              1461  
AAVDEFTSGVPIGEISRGGLIVSKIDVDG-KN-EYGSQADRRKMLMFGVGVLKFTVKPRAYSIERTLWLSDSVSPSL  
[: : :] [: : :] [: : :] [: : :] [: : :] [: : :] [: : :] [: : :] [: : :] [: : :] [: : :] [: : :]  
VATDALLKTTKDTILAMQIININVYFVDADNINTEHHKSTLFWSDFTGGVGLPLVYVKPFQDKYKRWLLRNSCAFTV

-661-

240 250 260 270 280 290 300

1491 1521 1551 1581 1611 1641 1671 1701

AMIREYDMIVDGDYLTITINSSEVNERGEKILDSIKASLGIL\*EVSFVLYSNRRPAYCVNRRNLOKHIDLLVFMIPDR

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32

KIVLEADERHGKTDLSMTABEALADKHKMKLDVYMDVADMVV

320 330 340 350

SEQ ID 8604 (GBS294) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 167 (lane 6 & 7; MW 65kDa – thioredoxin fusion), in Figure 238 (lane 2; MW 65kDa) and in Figure 40 (lane 6; MW 37kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 3; MW 76kDa).

Purified Thio-GBS294-His is shown in Figure 244, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 578

A DNA sequence (GBSx0618) was identified in *S. agalactiae* <SEQ ID 1807> which encodes the amino acid sequence <SEQ ID 1808>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

|          |                    |               |                    |
|----------|--------------------|---------------|--------------------|
| INTEGRAL | Likelihood = -3.61 | Transmembrane | 24 - 40 ( 20 - 41) |
| INTEGRAL | Likelihood = -1.97 | Transmembrane | 53 - 69 ( 52 - 72) |

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.2444 (Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB60012 GB:U09422 ORF21 [Enterococcus faecalis]  
Identities = 136/473 (28%), Positives = 228/473 (47%), Gaps = 40/473 (8%)

Query: 9 RGIKVKPYMRYSYL-PSFLFILFLTPVGVYSYYVLDL-----LKKKKKMSM---I 56  
RG +++P + + + + L + FL VG+ + L DK+ + I

Sbjct: 4 RGRIRPGSGKDLVPHFTIASILLVFLVLVGLPHVKTIQINWQPFSLQAKKIDIPILII 63

Query: 57 SVGTGLFLAFFVWYLTWFLQEBANPLPKLDRKMRSKFLYENGVYVEKR-----KKS 109  
S + + V++ F + +L +++K + EN + ++ K S

Sbjct: 64 SFSVAILLCILLVAFV---PKRVRYITVQKILRHQKLAEMILENGWYSEBQVKTEGPFKDS 120

Query: 110 NKIKTKTKYR-FPKVYVKQGYKDYLSVSFEMAQGRKPKKPKDIOGSELDTFPFMDPKMETDDP 168  
+TK K FPK+Y + + + E+ GK+Q + + +LE + + +K

Sbjct: 121 AGRTKEKITYPFPMYRYLNGLIQIRVEITLKGQDQLLHLEKLSGLYCELDTKELKD 180

Query: 169 RPKIKLAYSAFLSRITVKDVIWVKDEGIKIMDGYWDFINDPHLVAGTGGGRTVLLR 228  
+ Y L Y SRI++ D + KD +IM +W++ PH+L+AGTGGGRT

Sbjct: 181 SYVEYTLTYTTIASRISI-DEVEAKDGKRLMKRVWWSYDKLPHMLINGTGGGRTYFIL 239

Query: 229 SILRCLAEI-GVCDICDPRKADFTVMSDLGAFEGRIAPEKADIIEKFENAVTIFARYDF 287  
+++ L I DPK AD ++DL + + K D+ + E M R +

Sbjct: 240 TLIELLHTSDKLYILDPKNAD---LADLSGVANVTYRKESDLSLCISTFYESMKRSE- 295

Query: 288 VRNENKRLGKDKMKFYDY-GLRPFYFVCEYNALMSSLSYOEREIVDNPATFOYILLGRQ 346  
ENK+++ + K Y Y GL +F + DRY A M L +E V N Q ++LGRQ

Sbjct: 296 ---ENKMGNYTKGNYAILGLPAHLIFDEYVAFMEMLGTKEHTAVNKKLQIVMLGRQ 352

Query: 347 VGCNAILIAMQKPSADDLPTKIRSNMHHSVGRLLDGGYVM4PFDENENKRFPRKILAG 406  
G I+A Q+P A L IR +++GR+ + GY MMFG + + K+P F+K

-662-

Sbjct: 353 AGFFLILACORPDAKYLGDGIRDOFNFRVALGRMSENGYGMFGSDVO-KDF-FLK---- 406

Query: 407 RRVYGRGYSAVFGEVAREFYSPLLPKNFSFYDAFEKINRHENPFDPTENQEV 459

Sp|ct: 407 -RIKGRGYVDVGTSVISSEFYTPLVPKGYDFLEIKKLSMSRSTOATCRAEVA 458

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8605> and protein <SEQ ID 8606> were also identified. Analysis of this protein sequence reveals the following:

```

LipoP Possible site: -1 Crend: 8
MoG: Discrim Score: -10.05
GVH: Signal Score (-7.5): -3.42
 Possible site: 40
>>> Seems to have no N-terminal signal sequence
ALOM program count: 2 value: -3.61 threshold: 0.0
INTEGRAL Likelihood = -3.61 Transmembrane 24 - 40 (20 - 41)
INTEGRAL Likelihood = -1.97 Transmembrane 53 - 69 (52 - 72)
PERIPHERAL Likelihood = 1.01 224
modified ALOM score: 1.22

```

\*\*\* Reasoning Step: 3

```

----- Final Results -----
 bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

29.9/52.7% over 456aa

EGAD|17035| hypothetical protein Insert characterized  
GP|532554|qb|AAB60012.1|U09422 ORF21 Insert characterized

Enterococcus faecalis

```

ORF00100 (319 - 1677 of 2316)
EQAD|17035|17250|2 - 458 of 461| hypothetical protein {Enterococcus faecalis}
GP|532554|gb|AB60012.1||U09422 ORF21 {Enterococcus faecalis}
%Match = 11.2
%Identity = 29.9 %Similarity = 52.7
Matches = 135 Mismatches = 199 Conservative Sub.s = 103

```

FQVVCLKPIHHHLKRKLQMETHQNKKHLLINKR\*RGNLARLIPQYRGKVPMRYMSGYIL-FSFLFILFLTPTGVV  
:  
MKQRCKRIKPSGKDLPVFHTIASLLPVFFLVVLGL

10                  20                  30

426            453            483            513            570            600  
Y-----SYYYLD-LKKMMKMSISVGTGLFAFFVSYYLTWFLQGAN-PLFNKDLRLKRLSKMFKLYENGYYVYK-----  
          :  :  |  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
FHVKTIQQINWQDFNLQAQDKDIPYLLISFSAVAILCLLVAFFVFRVRYTQKQLYHRQKILAKMLLENKW-YESEQVKT

[illegible]

[illegible]

SEQ ID 8606 (GBS216) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 3; MW 66.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 579

A DNA sequence (GBSx0619) was identified in *S. galactiae* <SEQ ID 1809> which encodes the amino acid sequence <SEQ ID 1810>. Analysis of this protein sequence reveals the following:

Possible site: 14  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 580

A DNA sequence (GBSx0620) was identified in *S. agalactiae* <SEQ ID 1811> which encodes the amino acid sequence <SEQ ID 1812>. Analysis of this protein sequence reveals the following:

Possible site: 28  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.0944(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10219> which encodes amino acid sequence <SEQ ID 10220> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 5 Example 581

A DNA sequence (GBSx0621) was identified in *S.agalactiae* <SEQ ID 1813> which encodes the amino acid sequence <SEQ ID 1814>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence  
 10 INTEGRAL Likelihood = -6.94 Transmembrane 810 - 826 ( 808 - 830)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2975(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

IGB:D90354 surface protein antigen precursor [Strept...  
 20 >GP:BAAL4368 GB:D90354 surface protein antigen precursor  
 [Streptococcus sobrinus]  
 Identities = 151/408 (37%), Positives = 219/408 (53%), Gaps = 27/408 (6%)  
 25 Query: 451 PSKAVIDEAGQSVNGKTVLPWAEINIVAKQDFSGYKMTASQGKIANKVFFIDYKDDAL 510  
 P K +E G ++GK+VL Y D QYNG +++ I K F ++DIY ++AI  
 Sbjct: 1162 PHKVNKNMGVVDIGKSVLGGTITTYELVLDQYKGRSAKETIQGFFVVDYIPEAL 1221  
 30 Query: 511 DGKSMKNSIKASDGTDSVQL-LEMHVLSSTDILDEKLQTLIKANGISPGKPTMMITAKD 569  
 D ++ + IK +D + + + S + +Q ++K+A I+P G F ++TA D  
 Sbjct: 1222 D---LRITDLIKLITDANGKAVTVGSVADYASLEAAPAAVQDMLKXANITPKGAPQVFIADD 1278  
 35 Query: 570 PQAPFYKAVVQKGLDVTYNLSFKVKKEFTK--QIQNGVAQIDPGNGYTGNIIVADLITPE 627  
 PQAPFY AVY G D+T VK E K G +N QIDPGNGY NIV+N+  
 Sbjct: 1279 PQAPFYDAYVVTGDLTIVIPMTVKAEMGKIGSSYENKAYQIDPGNGYBSNIVINVPQIN 1338  
 40 Query: 628 IHKDV---LDEKDGKSNIGTQVLKGDVTVKLEGMVVTGSGYDLFEYKFDVLQKRPDL 684  
 KDV +D D +++ T+ L Y+L G ++P + +LFEY F D +T D  
 Sbjct: 1339 PEKDVTLTMCPADSTNVDCQTIALNQVFNRYLIGIIPADHAELEFSPSSDQDQDQ 1398  
 45 Query: 685 YLRD-KVAVKVDVTLKDGTVIKKGITNLGSETVTVNKKTLGYLVFPKDFLEKVARSEF 743  
 Y K AKVD-TLKDGT+IK GT+L YTE ++ G + FK+DFL V+ S F  
 Sbjct: 1399 YTCQYKAFKVDLTLKDGTTIIRAGDILTSYTEAQVDEANQGIIVTFKEDFLSVDGSAP 1458  
 50 Query: 744 GADDFVUVVKRIKAGDVYNIAEDFFINCKVKETSTVTHITE--KPKPVREQ----- 791  
 A+ ++ +KRI G NT +NG +TV T TPE +P PV+P  
 Sbjct: 1459 QAEVYLCKRIAGVITFANTYVNTVAGITYSSNTVRTSTPEPKQSPVDPKTTITTVFQPR 1518  
 Query: 792 --KATPKAPAG--LPQIGESAVAPLTAIGAILGA-IGLAGFKKRRK 834  
 KA AP G LP TG++S A L LG +L+A LG +++++  
 55 Sbjct: 1519 QGKAYQFAPPAGAQQLPATGDSNAYLPLGLHLSLTAGFSLGLRRKQD 1566  
 Identities = 75/242 (30%), Positives = 120/242 (48%), Gaps = 33/242 (13%)  
 Query: 11 SADQVTTQATTTQVTCNQAEVTVSTQLDKAVATAKKAANAVITTAIVNHAHTTIAQADLA 70  
 S+ T+QA T + V++++LD+A ++A+V V+ A VN T + D A  
 60 Sbjct: 73 SSQAETSQAQAGQKTKGMSVDVSTSEIDEAASKAQAGVTVSQDATTNKGTVETS--DEA 130  
 Query: 71 NQOT-VKDVTAQAQNTQAKDATNAENKIDAEKASQSVQSLNQTAKRID---AEN 126  
 NQ +T +KD +K A+ I+ T+ A N+AE+ R+Q NA KA+ A N  
 65 Sbjct: 131 NQKETEIKDDYSKQAD---IQKTEDYKAAVANKQAEIRITQENAAKQAEYQDLAAN 187

-665-

Query: 127 KDAQAKADATNAQLQKDYQAKLAKI KSVYAYNGVVRORNKDAQA-----KA 172  
 K + NQK + DK+AKLA+ + A V+Q N D+QA +  
 Sbjct: 188 KAEVERITNENQAKADYEAKLAQYQKDLA---AVQANNDQAAYAAKAYDKELARV 244

5 Query: 173 DATNAQLQKDYQAKLA---LYNQALAKAKABADKQGINNVAFDIKAQ---AKGVYNAHYG 225  
 A NA +K+Y+ LA N+ +KA+ A +Q D +A+ K + A +G  
 Sbjct: 245 QANNAAKKEYEALAAANTTYKNEQIKAEENAIQQRNAQAKADYEAKLAQYKDLAAQSG 304

10 Query: 226 NS 227  
 N+  
 Sbjct: 305 NA 306  
 Identities = 63/223 (28%), Positives = 100/223 (44%), Gaps = 31/223 (13%)

15 Query: 2 ITTLQTSQVADQVTTQITTTQVTONQAETVISTOLDKAVATAK-----KAVA 50  
 + +Q + +A + +A T+N+ + + + A AK K A  
 Sbjct: 241 LARVQANAAAKKEYEALAAANTTYKNEQIKAEENAIQQRNAQAKADYEAKLAQYKDLAA 300

20 Query: 51 VITTAIVNHAITTDQADLANQTVTVKDVATAK-QANTQAIKDATAENAKIDAENKAEQ 109  
 + A N A +A + V+ A A CR QA+ TA+NA+I AEN+A Q  
 Sbjct: 301 AQSGNATNEADYQAKAAEYEQELARVQANAAAKQAYEQALAAANTTAENAIQQ 360

25 Query: 110 RVSGNATQYKAKIDAENKDAQAKADATNAQLQKDYQAKLA---KIKSVYAYNGVVRORN 165  
 R +Q A +AK+ KD A A + NA + DVQ KLA ++ V+A NA +Q  
 Sbjct: 361 RNQAKNVEAKLAQYQKDL-AAAQSGNATNEADYQEKLAAYEKEELARVQANAAAKQEV 419

30 Query: 166 KDAQAKADATNAQL-----QKDYQAKLALYNQAL 194  
 + +A+A NA++ + DY+ KL+ Y + L  
 Sbjct: 420 BQKVOENAKVAEITTEANRAIRERNKAKTDYELKLSKYOEEL 462  
 Identities = 75/243 (30%), Positives = 101/243 (40%), Gaps = 56/243 (23%)

35 Query: 8 SQVSEAD-QVTTQITTTQVTONQAETVISTOLDKAVATAKKAQAVVTTAAVNHATTITDAQ 66  
 S+ +AD Q TT+ V MORET TQ + A A+ A V T +AQ  
 Sbjct: 142 SKQADIQKTTEDYKAVANQAETDRITQ-ENAAKQAYEQDLAANKAEVERTITNEAQ 200

40 Query: 67 ADL---ANQOTQVVDVTAQAQNT-----QAIK 91  
 A A Q KD+ A QAN +A+  
 Sbjct: 201 AKADYEAKLAQYQKDLAAVQANNDQAAYAAKAYDKELARVQANAAAKKEYEALAA 260

45 Query: 92 DATAENAKIDAENKARSQRVSQALNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA-- 149  
 T +N +I AEN A CR +Q A +AK+ KD A A + NA + DYQAK A  
 Sbjct: 261 ANTTNBOIKAEENAIQQRNAQAKADYEAKLAQYKDL-AAAQSGNATNEADYQAKKAY 319

Query: 150 --KIKSVYAYNGVVRORNKDAQAKADATNAQL-----QKDYQAKLALYNQAL 193  
 ++ V+A NA +Q + A A A NAQ+ + +Y+AKLA Y +  
 Sbjct: 320 EQELARVQANAAAKQAYEQALAAANTTAENAIQQRNAQAKADYEAKLAQYQKDL 379

Query: 194 LGA 196  
 L A  
 Sbjct: 380 LAA 382

There is also homology to SEQ ID 598.

SEQ ID 1814 (GBS191) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 582

A DNA sequence (GBSx0622) was identified in *S. agalactiae* <SEQ ID 1815> which encodes the amino acid sequence <SEQ ID 1816>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

-666-

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10221> which encodes amino acid sequence <SEQ ID 10222> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9921> which encodes amino acid sequence <SEQ ID 9922> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

15                   >GP:AC92523 GB:AF027768 TbpA [Serratia marcescens]  
                   Identities = 168/385 (43%), Positives = 232/385 (59%), Gaps = 13/385 (3%)

Query: 26 MHFQVAVGPPPERPCPGFD-KLYKSSRPQLIMDLPIRLKKVGLHMRKRYKCRBOGST 84  
           N F+V+ V P C BCG + + E+ DLPI KRV L + RRY CR C +T  
 20    Sbjct: 1 MHFQVD-VDPPLACBQGVQGEFVRPGKRDVFPDLPIHGKRVTLAVVRRYTCRACKTT 59

Query: 85 IS-----VDEKPSMTKLLKSIQSQSMKTFVEVAHSVGVDEKTIHVFKDYVALKERE 138  
           VD R MT RL + +++S + + VA G+DEKT+R++F R  
 20    Sbjct: 60 FRPQLPEAVDGR-NTLRLEHYVEKESFNHPYTFVAAQTGLDEKTVADIFNRAEFLGRN 118

Query: 139 YQETPHNLGIDEIHLIIRFPRVLNINERTIYDIKPNRNKETVIQPLSEISURTYIEYV 198  
           ++FETP+ LGIDE+++ +R R +LINIE RT+ D+ R ++ V L++ DR +EV  
 25    Sbjct: 119 HRFETPRILGIDELYNKRYRCILNIEERTLDELATRRQDVVINYLMLEDRQKVEIV 178

Query: 199 TCMHWKPKYDAVNTILPQARVVDKPHVVRMANQALDNVRKSLKAHMSQKERTLARERF 258  
           +MEMN PY+ AV +LPQAR++VVDKPHVVRMAN AL+ VRK L+ + + RTL +R  
 30    Sbjct: 179 SMDMNFPYRAAVKAVLPQARIVVDKPHVVRMANADALERVRKGLRKLKPSQRTLKGRK 238

Query: 259 ILLKREKHDNLNERESFLDTWLGNLPAKRAYELKEEFVHWINDTPDPBGHLRAYSCWRHRC 318  
           ILLKR H++++RE +++TW G P L AYE KE FY IWD + +W  
 35    Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAPQLAAVEHKKERFYGINDATRLQAEALDEWI-AT 297

Query: 319 MSNSKDAYKDLVRAVDNMHVEIFNYF--DKRLINAYTESINSIIRQVERMGRGYSFDAL 376  
           + K+ +DLVRAV NW E YF D +INAYTESIN + + R GRGYSF+ +  
 40    Sbjct: 298 IPKGQKEVWSDLVRAVGNMREETHYFETDMPVTINAYTESINRLAKDNKRGSGYSFEVM 357

Query: 377 RAKILFNKKLHKRKPFRNGSAFNK 401  
           RA++L+ K HKK+ P S F K  
 40    Sbjct: 358 RARMLTYTK-HKKAPTAKVSPFYK 381

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 583

A DNA sequence (GBSx0623) was identified in *S.agalactiae* <SEQ ID 1817> which encodes the amino acid sequence <SEQ ID 1818>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55                   bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>



-667-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]  
 Identities = 411/546 (75%), Positives = 483/546 (88%)

Query: 1 MNKFKVNISGMTCTGCEKHVSALKKIGARNIESSYKRGAEVFLPDDIEVSAIKAIDE 60  
 M K++V++ GMTCTGCE+HV ALA +GA IE +RGEAEVFLP++ +VE+A KAI +  
 10 Sbjct: 1 MKKYRVVDVGMTCTGCEBHVVALNMGATISVDVFRGAEVFLPDAVGSTAKKAISD 60

Query: 61 ANYQAGRIEVSLENNVALINEDNYDLLIISGGAAPSSAIIAEYGAQVMIGRTVGG 120  
 A YQ G+ REV S E V L NE +YD +IISG AAPSSAI++YGAQV MIERGT+GG  
 15 Sbjct: 61 AKYQPGKAREVQSQBMVQLGNEGIDYVLIISGGAAPSSAIEAVKYGAQVAMIERGTGG 120

Query: 121 TCVNIGCVPSKTLRLAGEINHLSDNPFILQTSAGEVDLASLITQDKVLSELNRQKYM 180  
 TCVNIGCVPSKTLRLAGEINHL+K+NPF+GL TSAGEVDIA LI QK++LV+ELRN KY+  
 20 Sbjct: 121 TCVNIGCVPSKTLRLAGEINHLAONNPFVGLHTSAGEVDLAPLIQKIDNVLTELRNSKV 180

Query: 181 DLIDENYFDLIGKEAKFVDASTVEVNGTKLSAKRFLIATGASPLQISGLEKMOYLTST 240  
 DLID+Y F+LI+GEAKFVD TVEVNG +SAKRFLIATGASP+ P GL ++DYLTST  
 25 Sbjct: 181 DLIDYGFELIGKEAKFVDENTVEVNGAPISAKRFLIATGASPAKFNIPGLNEVDYLTST 240

Query: 241 LLELKKIKPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLIKEYDPEISSVSKAL 300  
 +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLIKEYDPEISSVSK+L  
 25 Sbjct: 241 LLELKKIKPKRLTVIGSGYIGMELGQLFHHLGSEVTLIQRSERLIKEYDPEISSVSKSL 300

Query: 301 IEQGINLVKATFERVEQSSEIKRVIVTVNGSREVIRSDQLLVATORGPNTDNLNLAAG 360  
 +EQGINLVKAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATOR PNT +LNL ANG  
 30 Sbjct: 301 VEQGINLVKATFYERIEQNSDGIKKHVEVNGKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGRNHEILINDPQOTSNEKIYAAGDVTLPQCFVYVAAVEGGIITDVAIGLAKKIDLS 420  
 VE G EI+I+D+ +T+N +TYAGDVTLPQCFVYVAAV+GS+ NAIGLAKK++L  
 35 Sbjct: 361 VEIGSRGEIINDYSRTNWRTYAAGDVTLPQCFVYVAAVQGGVAAPNAIGSLKKKLELS 420

Query: 421 VVPVTFPTMTVAIVGLTEBQAKBKGVD/KTSVLELGAVPRAIVNRETTGVFKLVADAE 480  
 VVP VTPT P +ATVGLTE+QAKB GY+VKSULEL AVRA+VNRETTGVFKLVAD++T  
 40 Sbjct: 421 VVPGVTFPTAPATVGLTEQAKBNGYEVKTSVLELGAVPRAIVNRETTGVFKLVADGKT 480

Query: 481 LKVLGVHVSERAGDVITYASLAVKFGILTIEDLTETLAFYLTMAEGLKALVTFDKDISK 540  
 +KVLG H+V+ENAGDVITYAR+LAVKFGILT++D+ ETLAFYLTMAEGLKALVTFDKDISK  
 45 Sbjct: 481 MKVLGRHVVAENAGDVITYATLAVKFGILTVDIDRETALFYLTMAEGLKALVTFDKDISK 540

Query: 541 LSCCAG 546  
 LSCCAG  
 45 Sbjct: 541 LSCCAG 546

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 584

A DNA sequence (GBSx0624) was identified in *S.galactiae* <SEQ ID 1821> which encodes the amino acid sequence <SEQ ID 1822>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

55 Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4529 (Affirmative) < succ>  
 60 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-668-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA83973 GB:AF138877 mercury resistance operon negative  
regulator MerR1 [Bacillus sp. RC607]  
Identities = 84/129 (65%), Positives = 105/129 (81%)

Query: 1 MIYRISEFADKCGVKNKETIRYRYERKNLLQPHRTEAGYRIYSYDDVKVGVFKIRIQELGF 60  
M -R I E ADKCGVKNKETIRYRYER L+ EP RTE GYR+YS V R+ FIKR+QRIQELGF  
Sbjct: 1 MKFRIQELADKCGVKNKETIRYRYERGLIPEPEPRTSEGYRMYSQQTVDRLHPIKRMQELGF 60

Query: 61 SLSEITYLLGVVDKDEVRCDQMFVSKKQKEVQKQIEDLARIETIMDLKQRCPEDEKL 120  
+L+EI KLGVDV+DE +C+DM++F K +++Q++IEDLARIETIM DLK+RCP+ K +  
Sbjct: 61 TLNEIDKLLGVVDREAKRCDMYDFTLLKIEDIQRIEDLARIETIMDLKRCPEKNDI 120

Query: 121 HSCPIIETL 129  
+ CPIIETL  
Sbjct: 121 YECPIIETL 129

20 There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 585**

A DNA sequence (GBSx0625) was identified in *S. agalactiae* <SEQ ID 1823> which encodes the amino acid sequence <SEQ ID 1824>. This protein is predicted to be Nramp metal ion transporter. Analysis of this protein sequence reveals the following:

Possible site: 53  
>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -13.85 Transmembrane 175 - 191 ( 169 - 201)  
INTEGRAL Likelihood = -11.94 Transmembrane 150 - 166 ( 132 - 173)  
INTEGRAL Likelihood = -9.45 Transmembrane 491 - 507 ( 481 - 509)  
INTEGRAL Likelihood = -8.92 Transmembrane 375 - 391 ( 374 - 396)  
INTEGRAL Likelihood = -8.39 Transmembrane 72 - 88 ( 69 - 93)  
INTEGRAL Likelihood = -7.96 Transmembrane 280 - 296 ( 274 - 299)  
35 INTEGRAL Likelihood = -7.17 Transmembrane 413 - 429 ( 411 - 431)  
INTEGRAL Likelihood = -6.79 Transmembrane 327 - 343 ( 322 - 346)  
INTEGRAL Likelihood = -3.40 Transmembrane 444 - 460 ( 443 - 462)  
INTEGRAL Likelihood = -3.24 Transmembrane 132 - 148 ( 132 - 149)  
40 INTEGRAL Likelihood = -0.96 Transmembrane 115 - 131 ( 114 - 131)

----- Final Results -----  
bacterial membrane --- Certainty=0.6540 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF03825 GB:AE003939 manganese transport protein [Xylella  
fastidiosa]  
Identities = 185/450 (41%), Positives = 278/450 (61%), Gaps = 29/450 (6%)

50 Query: 16 ANGSGLEEINOTIEVFKDLSEFKTLIAYSGPGALVAVGYMDPGNNSTSTGGQNFQYLLI 75  
++ PSL E++ ++ V + + LLA+ CPG +V+VGYMDPGNN+T + GG F Y+L+  
Sbjct: 35 SDSFSLGEMIASVAVSRGHGFPRLAFLGPGYVSVGYMDPGNNATGLAGSGSRFGYMLL 94

55 Query: 76 SIILMSLIAMLLQMSARKIGIVTQMQLAQAIRARTSKQLGIVLMILITELAINATDIARY 135  
S+IL+S++++A+LQ +A+LGI + MDLAQA RAR S+ + LN++ ELAI+A D+AEV  
Sbjct: 95 SVILLNENVAIVLQALAAELGLIASDMDLAQACRAYSRSQTTLAVLVVCELAIAIADIAVY 154

Query: 136 IGGAIAYLLPLHPIAIAVFTTVFDVLLLLLTXTIGFKRIEAVVALLIWLIFVAYQA 195

-669-

IG AIAL LL +P+ V IT DV+L+LLL GFR +KA V+AL+VWIF F Q+

5 Sbjct: 155 IGTALALNLLGLVPIIIVGVITAVDVVLLMLHMRGFRALFAVIALLLVPGCFVQIV 214

Query: 196 LSHPIINTDIFKGLVPTSAFSTSHTVKQTPLSGALGIIGATVMPFNLYLHSSVQSRKL 255

5 L+ P ++ G VP + V L A+GI+GATVMPFNLYLHSS+VQ+R

Sbjct: 215 LAAPFLQEVLLGGFVFWQ-----VVADPQALYLAIGIVGATVMPFNLYLHSSVITQRAY 268

Query: 256 DHNNKKDIAR--AIRFSTFDSNIQLTVAFVFNSELLIMGVAVFKTGSVTPSPFFGLFKAL 313

10 + + R A+R++ DS + L +A F+H+ +LI+ AVF D

Sbjct: 269 P---RTPVGRRSALRWAVADSTLALMLALFINASTILITAAVFAHQHFD----- 315

Query: 314 SNSTIMSNILAIHASSGILSLLFAIALLASQNSTITGTLTGQIIMBGFHMKVPWIFR 373

15 + +LA + G+ +LFA ALLASG NST+T TL QQI+MBGF+ +++ W R

Sbjct: 316 VERIEQAVQLLAFVLGVGAATLFTALALASGINSTVTATLAGQIVMBGFRLRLRPLWR 375

Query: 374 RIITRLISVIPVMVLICVLTSGRSTVERHIAINMLNNSQVFLAFALFPMLELLIPTNSK 433

20 R++R +++++PV++ V + + T L+ SQV L+ LFP+++PLL +

Sbjct: 376 RVLRGLGLAIVPVIIVVALVGEQGT-----GRLLLSQVILSMQLPFAVPLRLCAVR 428

Query: 434 VEMDDDFKFWIILKILWLSVIGLIYLNK 463

25 M W++ ++ WL ++ LN+K

Sbjct: 429 KVMGALVAPRWLM-VVWMLNGVIVVLNK 457

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 586

A DNA sequence (GBSx0626) was identified in *S.agalactiae* <SEQ ID 1825> which encodes the amino acid sequence <SEQ ID 1826>. Analysis of this protein sequence reveals the following:

30 Possible site: 20  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2590 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 587

A DNA sequence (GBSx0627) was identified in *S.agalactiae* <SEQ ID 1827> which encodes the amino acid sequence <SEQ ID 1828>. Analysis of this protein sequence reveals the following:

45 Possible site: 53  
>>> Seems to have no N-terminal signal sequence

|             |                    |               |                        |
|-------------|--------------------|---------------|------------------------|
| INTEGRAL    | Likelihood = -9.82 | Transmembrane | 212 - 228 ( 204 - 233) |
| INTEGRAL    | Likelihood = -8.39 | Transmembrane | 98 - 114 ( 94 - 125)   |
| INTEGRAL    | Likelihood = -7.22 | Transmembrane | 132 - 148 ( 122 - 154) |
| 50 INTEGRAL | Likelihood = -6.42 | Transmembrane | 159 - 175 ( 155 - 188) |
| INTEGRAL    | Likelihood = -4.76 | Transmembrane | 54 - 70 ( 51 - 72)     |
| INTEGRAL    | Likelihood = -2.97 | Transmembrane | 18 - 34 ( 15 - 36)     |

----- Final Results -----

-670-

bacterial membrane --- Certainty=0.4927 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB16051 GB:Z99124 yydJ [Bacillus subtilis]  
 Identities = 97/239 (40%), Positives = 154/239 (63%), Gaps = 3/239 (1%)

10 Query: 4 LEFRKSIRGRITLFYIISTVALTYVLGYILFVGIDKIRHLTLGEPYPSITYTVFTQGFILIF 63  
 LEF+KSI + + + + +LG Y L V G I D K + + + T F + S + Y T V T Q G F + + F  
 Sbjet: 3 LEFRKKSISNKVIIILGAMFVFLFLGLYFLILVGDVDRKSVNTPTEMFSSYTVATQGLMLF 62

15 Query: 64 GFVIVVFFNKKDYSKCILYHYFSGYHLTKYFYTKLLVLFSEFFIALIVCNILASLLWGS 123  
 FVI +F N++YS+K IL++ G++ +FY K+ VLF E F I + ++ SL++ +  
 Sbjet: 63 SFVIAFFINREYSKNILFYKLIGENITYTFYKIAVLFLBCFAPITLGLLIISLMTY-HD 121

20 Query: 124 LFFVLTITLFLSLAVLQYLLVSTISILFSSNINSGVTIPFWITSLILWAIGG-IFKVS 182  
 +F LFS V+LQY+L++ TIS+L N+L+SIGV+I TW+TS+ILVAL F  
 Sbjet: 122 FSHALLLFLFSAVILQYLLIIGTISVLCPNLLISIGSVITWMTSVILVALSNKTPGFI 181

Query: 183 AIFDASNSLYKIKG-LFSPHWITLDTDFVIFVPMICLSVISPLCLSNRKWLINMG 240  
 A F+A N++Y I + L S MT+ D I+ Y++ +I+ +++ S RW+ G+  
 Sbjet: 182 APFAGNTMYFRIERVLQSDNMTLGSNDVLFITLYLVSIITINATVLRFSKTRWIKGL 240

- 25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 588

A DNA sequence (GBSx0628) was identified in *S.agalactiae* <SEQ ID 1829> which encodes the amino acid sequence <SEQ ID 1830>. This protein is predicted to be antibiotic epidermin immunity protein F. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2901 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 40 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB16052 GB:Z99124 similar to ABC transporter (ATP-binding  
 protein) [Bacillus subtilis]  
 Identities = 100/209 (47%), Positives = 150/209 (70%), Gaps = 4/209 (1%)

45 Query: 1 MFINNYTLKIGNRILLENTNLDPFEGEINHLGRNGSGKSLAKDFIINRGYSPNDIYB 60  
 M I NYTLK+ + LL+T+L F G+INH+G+NG GRSLAKDF+ +N DI +  
 Sbjet: 1 MNTANYTLKVGKTKLLQDTLHFSSGKINHVGVKGVKSLAKDFILNNKSRIGRDIRQ 60

50 Query: 61 DTILISSYSNLPSTVDT----INDLERTIPWKLSEKIYOLLNINQISKTIVKLQLEDGQKQ 116  
 + +ISS SN+P+DV+ ++ L + K+ +I LNN++ I V +K LSDGQKQ  
 Sbjet: 61 NVSLISSSSNIPNDVSKDFILFLSKKFDKMDIKIAYLLNLDNLDGKILNLSGQKQ 120

Query: 117 KVKLVLVSLDKHIIILDEITNALDKKSVDENVFLQNYIQYPEKIIINISHDINNIRS 176  
 K+KLL L DK+II+LDEITN+LDK+V EI+ FL YIQ PEKIIINI+HD++++  
 55 Sbjet: 121 KKLKSLFLDEKNTIVLDEITNSLDKKTVEINGFLPNKYIQNPEKIIINITHDLSLKA 180

Query: 177 LKGNVFLIDNGKICKVDTLDDAISWYLGE 205  
 ++G+Y++ ++Q+I + +D I Y+ E  
 Sbjet: 181 IEGDYIYFHQGEIQQYHSDVKLIEVYINS 209

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1831> which encodes the amino acid sequence <SEQ ID 1832>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2760(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 49/174 (28%), Positives = 82/174 (46%), Gaps = 27/174 (15%)

```

15 Query: 3 INNYTLKIGNKILLENINLDFEEGEINHLGRNGSGKSLAK-----DFINRNGN 52
 I N G R + L N N++ +G++ L+G NG+GKS + K II G
 Sbjct: 23 IQNLKKSYGKRKTILNNVNMIFPKGKVALIGMGAGKSTIMKILTLGLVSKTSGSIIFGR 82

20 Query: 53 YFS-----NDIYEOTLI---ISYSNLPSPDVTINDL-ERTIPWKLSEKETYQLNINQI 101
 +S I E+ + +S+Y N+ T+ + E TI L+K + + I
 Sbjct: 83 EWSRRDLKIGSIIIEEPFLYKMLSAYDNMKVVTMLGVSESTILPLLNK-----VGLGNI 137

 Query: 102 SETVKLQQLSDGQKQKVLVLVLLSLDKHIIILDEITNALDKSVDIEINFLQNY 155
 K +KQ S G RQ++ + + L ++ILDE TN LD + E+ +++
25 Sbjct: 138 DKR-FVQFSLGMRQLGIALSLINSFKLLILDEPTNLDEITGIELREITIEF 190

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 589

30 A DNA sequence (GBSx0629) was identified in *S.galactiae* <SEQ ID 1833> which encodes the amino acid sequence <SEQ ID 1834>. This protein is predicted to be aminoglycoside 6-adenylyltransferase. Analysis of this protein sequence reveals the following:

```

35 Possible site: 33
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1780(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA29839 GB:X06627 ORF (str) [Staphylococcus aureus]

Identities = 91/289 (31%), Positives = 146/289 (50%), Gaps = 14/289 (4%)

```

45 Query: 1 MRDEQEIYNLVNINIANQDKRIEAVLLNGSRANPNVPKDDPQYDVIYVFNFIDISDIN 60
 MR E+E NLV A Q ++ + L GSR N N+ KD PQYD F + IE + +
 Sbjct: 1 MRTEKEILNLVSGFAYQRSNVKKIALBSRSTINENKKDKPQYDFAFFVSDIEYFTHRES 60

50 Query: 61 YHKKPGDILINQKPNH---FRNKTEYNCFAYINQFQOLIRIDRLIKPEFLDYILDA-- 115
 + FG++L +QRP + P +Y ++Y+M F+D ++D+ LI + L Y D+
 Sbjct: 61 WLSLFGELLFIQKPEINELFPFDLYG-YSYIMYPKDGIMKDITLNLKOLNRYFSOSDG 119

 Query: 116 FSKVLLDKNKYLDYNFERSLSLYETQKSKDRINKILNEIYVWSTYVVKGIARNDIYSR 175
 K+L+DK N S Y K+ +E E NE + VSTYV KG+ R +I+++
55 Sbjct: 120 LVKILVDKCNLVTQIRIVPDSSNYWLKKPTREFYDCCNFPWSVTVAKGVFRREILFAL 179

 Query: 176 FMISNFIKNAFTLLKQKILISKELDSLFGKLDKDKILQYITDKD--QLKIFSNKSLKD 233
 +N ++ ++++ I + D S GK K I +Y+TDK+ LL F +

```

-672-

Subject: 180 DHFNNILRPELLRMISWYIGFNRGFD-PSLGKNYKFINKYLTDKFNMLLATPEMNGYRK 238

Query: 234 IEANLRFLIDETNQAKYISINRKLNLNGEYQAMKFMNIFLSNSYQN 282

Spict: 239 TYCSFKLCC---ELFKYY8-NKVSLGNYNYNPNYEKNIENFIRNNYEN 282

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8607> and protein <SEQ ID 8608> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 5
MoG: Discrim Score: -5.26
GVH: Signal Score (-7.5): -6.14
 Possible site: 33
>>> Seems to have no N-terminal signal sequence
ALOM program count: 0 value: 6.10 threshold: 0.0
PERIPHERAL Likelihood = 6.10 151
modified ALOM score: -1.72

```

\*\*\* Reasoning Step: 3

```

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1780(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

31.0/53.4% over 281aa

Staphylococcus aureus

EGAD|9462| streptomycin resistance protein Insert characterized

SP|P12055|STR\_STAAU STREPTOMYCIN RESISTANCE PROTEIN. Insert characterized

GP|46644|emb|CAA29839.1||X06627 ORF (str) Insert characterized

PIR|S00938|S00938 str protein - plasmid pS194 Insert characterized

ORF00399 (301 - 1146 of 1452)

EGAD|9462|9267(1 - 282 of 282) streptomycin resistance protein (Staphylococcus aureus)

|    |        |           |                                  |    |       |     |            |        |     |
|----|--------|-----------|----------------------------------|----|-------|-----|------------|--------|-----|
| SP | P12055 | STR STAAU | STREPTOMYCIN RESISTANCE PROTEIN. | GP | 46644 | emb | CAA29839.1 | X06627 | ORF |
|----|--------|-----------|----------------------------------|----|-------|-----|------------|--------|-----|

```
(str) {Staphylococcus aureus} PIR|S00938|S00938 str protein - Staphylococcus aureus plasmid
```

bS194

\$Match = 12.8

%Identity = 31.0    %Similarity = 53.4

Matches = 87 Mismatches = 125 Conservative Sub.s = 63

```

117 147 177 207 237 267 297 327
**LMTY*H*TVENIWNHNQLLRKI*N*ILGRKG*MSLI*VDYMLREKYKGNIKVLEXTW*YKVK*EVAIRDECEIYN
 || :|||
 MRTEKILN

```

357 387 417 447 477 507 558  
 LVLNIANQDKRIEAVLLNGSRANPVPKDDQDYDIVFVNFIEDIISDTNYYKKPGDILIMQKNEFR--NKTETNCF  
 ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 LVSEPAYQRSNVKIALSGRSTENIKKKKKQDYDPAFFVSDIEYFTHESWLSLGFGLLIQKPEDMSLFPDDLGVY

[illegible][illegible]

-673-

```

1056 1086 1116 1146 1176 1206 1236 1266
DETNQMAKYISIRKLNLNQGEYSQAMKFNITPLSNSTQNFN*YTCVKINRL*LSKLNYS*RPFRKIINNFGDK*WDSK
: : : : : : : :
5 ELFKYYSNKVS-----CLGNVNTFYETKNTETFTKNTYEN
 260 270 280

```

SEQ ID 1834 (GBS46) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 6; MW 34.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 3; MW 59.8kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 590

A DNA sequence (GBSx0630) was identified in *S.galactiae* <SEQ ID 1835> which encodes the amino acid sequence <SEQ ID 1836>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

---- Final Results ----
bacterial cytoplasm --- Certainty=0.1179(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 591

A DNA sequence (GBSx0631) was identified in *S.galactiae* <SEQ ID 1837> which encodes the amino acid sequence <SEQ ID 1838>. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.81 Transmembrane 177 - 193 (177 - 194)
INTEGRAL Likelihood = -0.27 Transmembrane 129 - 145 (129 - 145)

---- Final Results ----
bacterial membrane --- Certainty=0.2125(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8609> which encodes amino acid sequence <SEQ ID 8610> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 9
McG: Discrim Score: -19.59
GvH: Signal Score (-7.5): -4.49
Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALCO program count: 2 value: -2.81 threshold: 0.0
INTEGRAL Likelihood = -2.81 Transmembrane 172 - 188 (172 - 189)
INTEGRAL Likelihood = -0.27 Transmembrane 124 - 140 (124 - 140)
PERIPHERAL Likelihood = 8.01 30

```

-674-

modified ALOM score: 1.06

\*\*\* Reasoning Step: 3

5       ----- Final Results -----  
               bacterial membrane --- Certainty=0.2126(Affirmative) < succ>  
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10   The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 592**

15   A DNA sequence (GBSx0632) was identified in *S.agalactiae* <SEQ ID 1839> which encodes the amino acid sequence <SEQ ID 1840>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have an uncleavable N-term signal seq

20       ----- Final Results -----  
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25   A related GBS nucleic acid sequence <SEQ ID 10223> which encodes amino acid sequence <SEQ ID 10224> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB49414 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]  
 Identities = 29/86 (33%), Positives = 52/86 (59%), Gaps = 4/86 (4%)

30   Query: 14   TYYILLALFE--RAHYAIMQKQVESMSGGDVRIAGTNYGATENLLKQGWKSIKPSD--D 69  
               +Y ILL L E + HGZYAI +++EE+ G + G +Y ++ L K K ++ ++  
   Sbjct: 19   SYLLILLTINENKLAGHYAIRKRLEELTDGKLVPSBQALYSILMLKIKCYKLVDYNAEYVG 78

35   Query: 70   RRRKVVITITGTGKEIVE:STNRLRLK 95  
               R R + Y ITE GKE+++ +R++  
   Sbjct: 79   RVRRYQITRLGKVLDRKKEETREI 104

No corresponding DNA sequence was identified in *S.pyogenes*.

40   Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 593**

A DNA sequence (GBSx0633) was identified in *S.agalactiae* <SEQ ID 1841> which encodes the amino acid sequence <SEQ ID 1842>. Analysis of this protein sequence reveals the following:

45   Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

50       ----- Final Results -----  
               bacterial cytoplasm --- Certainty=0.0510(Affirmative) < succ>  
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>



A related GBS nucleic acid sequence <SEQ ID 10225> which encodes amino acid sequence <SEQ ID 10226> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

5 >GF:AF22299 GB:AF185571 putative N-acetyltransferase Camello 2
 [Homo sapiens]
 Identities = 32/110 (29%), Positives = 54/110 (49%), Gaps = 4/110 (3%)

10 Query: 67 IKMAEQDDIPQIENYQNRKGQ-FWIALENERKVGSIALLAIDOKTAVLKKFFTYPKRG 125
 + +A + D+ I Y + G FW+A ERVVG++ L +DD T K+ +
 Sbjct: 86 VDIALRTDMSDITKSYLSEGGSCFWVASSEKRVGTVGALFVDDFTLREKRLQLFHLSD 145

 Query: 126 NPVR---LGRKLPERFMI.FARASKFTRIVLTPEKEKSHFFPYHQGFQK 172
 N R + + L + FAR ++ +VLDT + + Y++ GFK+
15 Sbjct: 146 NEHRGGGLAKALVKTVLQFARDUGYSEVVLUTSNLQISAMSLYQSLGFKK 195

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 20 Example 594

A DNA sequence (GBSx0634) was identified in *S.galactiae* <SEQ ID 1843> which encodes the amino acid sequence <SEQ ID 1844>. Analysis of this protein sequence reveals the following:

```

 Possible site: 47
 >>> Seems to have a cleavable N-term signal seq.

25 INTEGRAL Likelihood =-11.94 Transmembrane 159 - 175 (151 - 180)
 INTEGRAL Likelihood =-11.62 Transmembrane 231 - 247 (225 - 251)
 INTEGRAL Likelihood = -9.98 Transmembrane 182 - 198 (177 - 203)
 INTEGRAL Likelihood = -7.11 Transmembrane 118 - 134 (106 - 136)
30 INTEGRAL Likelihood = -1.49 Transmembrane 74 - 90 (74 - 93)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10227> which encodes amino acid sequence <SEQ ID 10228> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

40 >GP:CAB15891 GB:E99123 yxlg [Bacillus subtilis]
 Identities = 42/188 (22%), Positives = 94/188 (49%), Gaps = 4/188 (2%)

 Query: 1 MKSLAVMLKKWMMNVRTYKVISILITCSIFGLIQLPLTAMMDIMA--GLPKKLGQAI 58
 MK + +L+KEW+E ++ K+I + I I G+ FLT MP+I+A G Lp ++ +
45 Sbjct: 1 MKVMALLQKWEIGSKSLTWLPLAMHIVGLITQFLTYYMPETLHAGNLPDGNKISF 60

 Query: 59 PEPTIDSYIQYKNNQGLGVILV7LPSSTLTQEPSEKGTLINLWPTKGLAKVLIILAKFI 118
 P+ + + N LG+ ++F ++ E ++G ++++ + I++K++
50 Sbjct: 61 TMSQSEVMVSTLQFNTLGNALVIPSVMGSHVNRQGVTLIMSRFVTAHTYVSKL 120

 Query: 119 VITLLNTVSYLLSVVHFSYLYFSENGSHKLMVYGATWFIGILFT--GLLPFSVLGRK 177
 + +++ +S+ + + Y F + + + + ++FT + L S +FR
 Sbjct: 121 IQSVIGIMSPAAGYGLAYTYVRLFLZDASPSRFAASGLVLYLWVIFIVTAGLASETIF- 179

 Query: 178 TLGGLLGC 185
 ++G C
55 Sbjct: 180 SVGAAAC 187

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 5 Example 595

A DNA sequence (GBSx0635) was identified in *S.agalactiae* <SEQ ID 1845> which encodes the amino acid sequence <SEQ ID 1846>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```

Possible site: 14
10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10229> which encodes amino acid sequence <SEQ ID 10230> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:CAB12736 GB:Z99108 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 105/299 (35%), Positives = 175/299 (58%), Gaps = 11/299 (3%)

Query: 4 ISPNVTKSPGPKILNNVSPDLNNMIYGVGPKGAGKTTTIEMILGLLKFDTGFTIF 63
25 + +NVTK+ + I+++SF + E ++GF+GNGAGKTTTI+M++GL+K G + I
 Sbjct: 5 LEKKNVTKNIRGRTIIDLSFTIRGEVFGFLGPKGAGKTTTIEMVGLAKLKGQVLIC 64

Query: 64 GKKNVGRDITNQLIGLYLDPVPEYYDMTALEYLDLCSGLARSKHKLSNKELLRSVGLDD 123
30 G+ + + IG + + PE Y +++ + L + + + K E+ VGL D
 Sbjct: 65 QGSITKEYAKAIKHGAIVENPELYKFLSGYKNLQGFARWVGVTPEKIDEVVGLVGLTD 124

Query: 124 N-HQKIATYSRGMKQRIQLGLAQALVHDPKILIIIDBPTSALDPKGRQDILDIISNLGEK-- 180
 H K+ TYS GM+QRLGLAQ L-HDPK++I DEPT+ LDP G ++I D + L E+
35 Sbjct: 125 RHEDKVTYSGLMRQLGLAQCLLHDPKVLILDPPTNGLDPAIRIRDLHKLTIRERGM 184

Query: 181 TVIPSTHILSDVEKICDHVLVLTCKGIYSLEELGKGSSENYSVRIKIKVTSEAKVLSH 240
 VI S+H+LQ++E +CD + +L K + ++K + +EN + ++ SEA + +
40 Sbjct: 185 AVIVSHELLSEMRLMCDRIALLQKGGKILIDQNVKNDENIDENITYFEQVR-QPSRAATVILN 243

Query: 241 NYQIEKKDNEYALTLEKSKMNKADLLAGFYQDLVSLKISPSAIEVIDNSLEHLYLEVTT 299
 Y + K N + L ++ +L LV +I ++VI SLE+ +LE+T
45 Sbjct: 244 QYDLKSTNGVEIKLAKKEVPAVIEL- - - - -LVMQQTIRIYEVKVTIKSLERDPLEMT 295

```

There is also homology to SEQ ID 686.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 596

A DNA sequence (GBSx0636) was identified in *S.agalactiae* <SEQ ID 1847> which encodes the amino acid sequence <SEQ ID 1848>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
50 >>> Seems to have no N-terminal signal sequence

```

-677-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4040 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAE71491 GB:U53767 ORF6 [Bacillus pumilus]  
 Identities = 39/134 (29%), Positives = 71/134 (52%), Gaps = 16/134 (11%)

10 Query: 2 LGENTYLQRTQIGMTQENLSDYLHLTKTTISKWENNQAQKPDIDYLILQANLPDISLDDLV 61  
 IG NI +R ++QE +++ L +++ ISKWE NQ++P +D LI +A LFD + +LV  
 Sbjct: 4 LGSNISRKRLKSLKLSQEVVAEQGVSRQAISKQVETNQSEPSMDNLRLAELPDSIDIKLV 63

15 Query: 62 GYQNTLSDDQGNQIKLIDKIKANVLSEKDFPQEVKELSKQFENDFKTLLIMNM--VLSN 119  
 S +Q ++ KDL+ + K++ Q F +L++I+ +  
 Sbjct: 64 -----SPEQYSEBQKDLLETRIE-----HQKDKINQMSAVFGRLIMLISFPGYIGA 109

Query: 120 LTNLADSENKESL 133  
 L +L+ ++ W L  
 20 Sbjct: 110 LFDLSSYQLPIWKL 123

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 597**

A DNA sequence (GBSx0637) was identified in *S. agalactiae* <SEQ ID 1849> which encodes the amino acid sequence <SEQ ID 1850>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have an uncleavable N-term signal seq

30 INTEGRAL Likelihood = -13.59 Transmembrane 152 - 168 ( 145 - 173)  
 INTEGRAL Likelihood = -9.71 Transmembrane 7 - 23 ( 3 - 27)  
 INTEGRAL Likelihood = -6.95 Transmembrane 125 - 141 ( 122 - 146)  
 INTEGRAL Likelihood = -4.51 Transmembrane 85 - 101 ( 83 - 102)  
 35 INTEGRAL Likelihood = -3.35 Transmembrane 55 - 71 ( 54 - 75)

----- Final Results -----

bacterial membrane --- Certainty=0.6434 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]  
 Identities = 51/186 (27%), Positives = 106/186 (56%), Gaps = 5/186 (2%)

45 Query: 5 SFFQCIVLLVSPVNLTLAVKSQSDMISYLDNITSAPFQSIHNPDLTNMTIISTVSPFLT 64  
 +F V+ L+ F + + S ++ + + + S Q +P LT+M + + S +  
 Sbjct: 10 APTIISVLISLIGFSRMAFTI-SANVYKPKRDVSI-LVQGWSPSLTIDIMKFTYIGSTAS 67

50 Query: 65 TSLIALVILGYQ-FLNQRIAWLWM-LFFGTINALALLKDIARHRP-MNQLVDFSGYS 121  
 +++LVIL + Y L R+ + LF + G+ L L+K R RP +++L+ GYS  
 Sbjct: 68 LILSLVILVFLYRILKRLRLVLFVAVMWGSLPLNMAVLFQFQARPDRLRIDIGYS 127

Query: 122 FPGHTISAFLLMILVLVVARQRRLRVLSQVVFVIFALVILASVIPSRLYLENPLTDLIL 181  
 FPGH ++AF L ++ + + + + + + +F+ + + + S+ SR+YL H+ +DI+  
 55 Sbjct: 128 FPGHAMMAVFAKGLITFLLRWHITARWARILLILFSLMLISLIGTSRIYLVGHYPSDII 187

Query: 182 GSLLAG 187  
 L G  
 Sbjct: 188 AGYLAG 193



-679-

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 5           bacterial cytoplasm --- Certainty=0.4288(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CA15470 GB:Z99121 yv0C [Bacillus subtilis]  
      Identities = 53/96 (55%), Positives = 70/96 (72%)  
      Query: 1 MDITDYQKWVSFEYKRNWYQVNSFIRENPLSEVGEAQAIRKYEIGRDRPQSTBQIDL 60  
              M + D +KN+ EFY+KR W +Y FIR FL EE GELA+A+R YEIGRDRPQE E +  
      15 Sbjct: 1 MQLADAKMKKEFYKRGWTEYGPPIRVGFMREAGSLARAVRAYEIGRDRPDEKESRA 60  
      Query: 61 ENLNDIKELGDLNDFILADQYNSLEETISAHR 96  
              E ++ EE+GDV+ NI ILAD Y +SLE+++ AH+  
      20 Sbjct: 61 EQKQELIEEMGDVIGNLAILADMYGSLVEDVMKQ 96

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 599

25 A DNA sequence (GBSx0639) was identified in *S.agalactiae* <SEQ ID 1855> which encodes the amino acid sequence <SEQ ID 1856>. Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.0635(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06803 GB:AF001517 unknown conserved protein [Bacillus halodurans]  
      Identities = 83/186 (44%), Positives = 117/186 (62%)  
      Query: 1 MRITIFCGASTGPNVYSEKTVLAQNMMAQKHSLVYGGCKVGLMGVMDTVIANGSYTT 60  
              M+I +FCG+S G + VY E L + +A+ +LVYGG VG+MG +AD+V+ GG  
      40 Sbjct: 1 MKLIVFCGSSNGASDVYKRGARQLGKELARRGITLVYGGASVGI MGAIVADSVLEAGGEVI 60  
      Query: 61 GVPIPTFLRDRETAHNLSELIIVNMMPERKAKMMLGDAFIALPGSPGTLEETISEVISWS 120  
              GV+P FL + EI+H +L++LI+V M ERKAK L D F+ALPGSPOTLEE E+ +W+  
      45 Sbjct: 61 GVMRFLREEPEISHPHLTKLIVVETIMHERKAKMAZLADGFIALPGSPOTLERFBEFTIWA 120  
      Query: 121 RIGQNDNP CILYNVNGYFNDLKNMFOHVMGEGFLSDRENVLFSDDITEIEDFITNYKV 180  
              +IG+ PC L N+N YF+ L + HM E FL + R L D + D + Y+  
      50 Sbjct: 121 QIGLHQKPCGLLNINHYFDPVTLTHMSENQPLHRYKYSMAVLVHTDPILLDCQSTYEP 180  
      Query: 181 PSTQRY 186  
              P+ + Y  
      Sbjct: 181 FTVKAY 186

55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-680-

**Example 600**

A DNA sequence (GBSx0640) was identified in *S. agalactiae* <SEQ ID 1857> which encodes the amino acid sequence <SEQ ID 1858>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -7.86 Transmembrane 222 - 238 (214 - 239)
 INTEGRAL Likelihood = -6.69 Transmembrane 39 - 55 (36 - 58)
 INTEGRAL Likelihood = -4.25 Transmembrane 266 - 282 (266 - 284)
 INTEGRAL Likelihood = -1.28 Transmembrane 166 - 182 (166 - 182)
10 INTEGRAL Likelihood = -1.01 Transmembrane 190 - 206 (190 - 206)
 INTEGRAL Likelihood = -0.96 Transmembrane 70 - 86 (70 - 86)

----- Final Results -----
 bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GF:CAB12420 GB:Z99107 ydIL [Bacillus subtilis]
Identities = 40/132 (30%), Positives = 63/132 (47%), Gaps = 8/132 (6%)

Query: 107 ESQNYDATFNI-----LMSYSVVVGPFEEVLYRGIVIANLL-SKYGNFAITSGILPG 160
 ES+N A ++ LMI S +VGP EE+++R I+ L K +FA +S ++FG
Sbjct: 114 ESENTQAILDVIQAVPLMIIVSSIVGFLEIIPRKIIIPGALYEKTNFFAGLISVIFG 173

Query: 161 LMEQDISQLLTSTIAGIIMGFI-AHYSPFKVALLMHICNPFIVEIPQLSTVNELYGTYP 219
 ++H D+ LL + G F+ A V + H+ N V+ QL V
Sbjct: 174 IVIADLKHLILLYTANGFTFAFLYARTERIRIWPVFAHLMNTEFV-VIMQLEPVRNYLEQQS 232

Query: 220 ENILLILAILFI 231
 + LI+ LF+
Sbjct: 233 TQWQLIIQGLFL 244

```

No corresponding DNA sequence was identified in *S. pyogenes*.

35 A related GBS gene <SEQ ID 8613> and protein <SEQ ID 8614> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 6
McG: Discrim Score: 12.52
GVH: Signal Score (-7.5): -1.74
40 Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 2 value: -6.69 threshold: 0.0
 INTEGRAL Likelihood = -6.69 Transmembrane 39 - 55 (36 - 58)
 INTEGRAL Likelihood = -0.96 Transmembrane 70 - 86 (70 - 86)
45 PERIPHERAL Likelihood = 4.56 21
modified ALOM score: 1.84

*** Reasoning Step: 3

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the databases:

```

Query: 10 LIGLILLQAIVLSLATTLEFASILLQNDVWIGIASTLIALLIPCF 53
 L+ L LL ++++LS++ +L +W+ ++A L+R ++ CF
Sbjct: 21 LLCLCLIVRSLLLSVSLYSALILLVLLVILNVTVATPLIAFVVSCEF 64

```

-681-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 601

A DNA sequence (GBSx0641) was identified in *S. agalactiae* <SEQ ID 1859> which encodes the amino acid sequence <SEQ ID 1860>. This protein is predicted to be capa protein. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -13.80 Transmembrane 27 - 43 (22 - 50)

----- Final Results -----
bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9385> which encodes amino acid sequence <SEQ ID 9386> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF13661 GB:AF188935 pX02-56 [Bacillus anthracis]
Identities = 68/224 (30%), Positives = 118/224 (52%), Gaps = 10/224 (4%)

Query: 95 FKEVKSWIESADLAIGDYBGTSISE----YPLAGYPL-FNAPNEIATTMKETGYDVVDLA 149
F+ V +++++D G++E + E Y A + +A E +KE G+ V++LA
Sbjct: 87 FRHVSPLYLNSDYVSGNFEHPVLEDKKYNQKADKNHLSAKETVKAKEAGFTVLMLA 146

Query: 150 HNHILDSQLAGAINVKTFFNRLGLDTIGVYLKORKNKEDILKHVNGIKIALISYSYGY-N 208
+NH+ D G +T+K F LD +G ++ ++I+ ++VNG+++A LG++ +
Sbjct: 147 NNHMTDYGAKGTDTIKAFKEADLYVGAGBNFKDVKNIVYQVNGURVATLGSPTDPAVA 206

Query: 209 GMEANVSKSDYBKMSDLDTKKIKQDIKKARKEADITIVMPQMGIEYKPKPTTBQVMLYH 268
G A + D+ K+I + + AD+ +V G EY KP+ Q L
Sbjct: 207 GATATKEQPGSLSMNPDVLLKQISKADPKKGNADLVVNTNHWGEYDNKPSPRQALAK 266

Query: 269 SMIKWAGADIIIPGGHPHVPESEVIKKDQKQKFIITYSMGNFISNQ 312
+M+ GADII G HPHV++ +V K+ I YS+GNF+ +Q
Sbjct: 267 AMVDAGADIIVGHHPVQLGSFVDYKQ-----IIFYSLGNFVFDQ 306

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1861> which encodes the amino acid sequence <SEQ ID 1862>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -12.05 Transmembrane 44 - 60 (40 - 68)

----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9119> which encodes the amino acid sequence <SEQ ID 9120>. Analysis of this protein sequence reveals the following:

```

Possible cleavage site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial membrane --- Certainty= 0.582(Affirmative) < succ>
bacterial outside --- Certainty= 0.000(Not Clear) < succ>

```

-682-

bacterial cytoplasm --- Certainty= 0.000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below:

Identities = 232/334 (69%), Positives = 273/334 (81%), Gaps = 4/334 (1%)

5 Query: 24 YQKTLIFCNVAVIIAIFILGLSKDLAQSGQKRVANNET---VKTRVAVANGDILLHDVLY 79  
Y+ET+ VA+I+A+ + GL DL + +A+ + VKTRVAVANGDIL+HD+LY  
Sbjct: 40 YKTMATVALIVALLLFGLTYDLILGVQKNEIAAQKSAQPKVKTRVAVANGDILLHDVLY 99

10 Query: 80 ASARQPDGTYNFTPYFKYKSWIESADLAIGDYGTISSEYPLAGYPLFNAPNEIATMK 139  
SAR+ D TY+FTPYF+ VK WI ADLAIGDYGTIS +YPLAGYPLFNAP EIA +K  
Sbjct: 100 MSARKADDTYDFTPYFYVKGWISGADLAIGDYGTISPDYPLAGYPLFNAPPEIAGALK 159

15 Query: 140 ETGYDVVDLAHNLHSDLSQLAGAINTVKTENKGLDITGVILKDRNKEDILIKHVINGKIA 199  
TGYDVVDLAHNLHSDLSQ LGA+NT K P+LG+D+IG+Y KDR+KE LIK+VINGKIA  
Sbjct: 160 NTGYDVVDLAHNLHSDLSQDGAINTKKVPHQLGIDISIGYDKDRSKESFLIKHVINGKIA 219

20 Query: 200 ILGYSYGYNGMBANVSKSDYKHMSDLDTKKIKQDIKKAKKADITIVMPQMGIEYQKKP 259  
ILGYSYGYNGMBA +8+ DYKHMSDLDT KIK+ +++ AEK+AD+ITVMPQMG EY +P  
Sbjct: 220 ILGYSYGYNGMBATLSQEDYKHMSDLDTBAKIKKELQLAKKADITIVMPQMGTEYALEP 279

25 Query: 260 TTEQWMLYHSMIKWGADIIFGGHFPVVEPSEVIKKGQKFFIYSMGNFISNQRLETVD 319  
T EQ LYH MI WGAD++ GGHFPV+EPSE + K QKKFFIYSMGNFISNQRLETVD  
Sbjct: 280 TAPQKLYHKMIDWGADVVLGGHFPVTEPSEVTKSRQKFFIYSMGNFISNQRLETVD 339

30 Query: 320 IWTREGLMDVTIRKKKGQKTVIKKKAHPTLVEA 353  
IWTREGLMD+T EKK KT IK V+AHPT+V A  
Sbjct: 340 IWTREGLMDLTFEKKDKNTKIKTVEAHPTVLA 373

30 A related GBS gene <SEQ ID 8615> and protein <SEQ ID 8616> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 7  
SRCLG: 0  
35 McS: Length of UR: 18  
Peak Value of UR: 3.83  
Net Charge of CR: 2  
McS: Discrim Score: 15.36  
GvH: Signal Score (-7.5): -1.52  
Possible site: 32  
40 >>> Seems to have a cleavable N-term signal seq.  
Amino Acid Composition: calculated from 33  
ALOM program count: 0 value: 4.35 threshold: 0.0  
PERIPHERAL Likelihood = 4.35 170  
modified ALOM score: -1.37  
45 \*\*\* Reasoning Step: 3  
Rule gp01  
50 ----- Final Results -----  
bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

30.6/53.3% over 230aa

Bacillus anthracis

60 HQAD|20151| capa protein Insert characterized  
SP|P19579|CAPA\_BACAN CAPA PROTEIN. Edit characterized  
GP|142633|gb|AA22288.1||M24150 46 kd encapsulation protein Capa Insert characterized  
PIR|C30091|C30091 capa protein - Insert characterized

ORF02075 (574 - 1257 of 1734)



```

BGAD|20151|20674|83 - 313 of 411) capA protein [Bacillus anthracis] SP|P19579|CAPA_BACAN
CAPA PROTEIN. QP|142633|gb|JAA22288.1|242150 46 Kd encapsulation protein CapA {Bacillus
anthracis} PIR|C30091|C30091 capA protein - Bacillus anthracis
%Match = 8.9
%Identity = 30.6 %Similarity = 53.3
Matches = 70 Mismatches = 102 Conservative Sub.s = 52

10
15
20
25
30

```

SEQ ID 8616 (GBS29) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 5; MW 40kDa), in Figure 181 (lane 6; MW 47kDa), in Figure 169 (lane 13 & 14; MW 54.5kDa – thioredoxin fusion) and in Figure 239 (lane 3; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 5; MW 65kDa).

SEQ ID 8616 (GBS289L) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 2; MW 72kDa) and in Figure 184 (lane 5; MW 72kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 5-7; MW 47kDa).

GBS289L-His was purified as shown in Figure 234, lane 9-10. Purified GBS289L-GST is shown in Figure 245, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45      **Example 602**

A DNA sequence (GBSx0642) was identified in *S.galactiae* <SEQ ID 1863> which encodes the amino acid sequence <SEQ ID 1864>. This protein is predicted to be thiamin biosynthesis protein ThiI (thiI). Analysis of this protein sequence reveals the following:

```

Possible site: 55
50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2720(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-684-

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 9971> which encodes amino acid sequence <SEQ ID 9972> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AA000308 GB:AF008220 YtbJ [Bacillus subtilis]
 Identities = 184/354 (51%), Positives = 249/354 (69%)

Query: 11 MCYSEIMIRYGRISTKKNNRMRPINKLKNMHEVLSIYPDVSVKITDRGRHVILNGTDYH 70
 M Y I+R+GB+STK KNR FI +LK N+ VL YP++ ++RDR + LAG D
Sbjct: 1 MNYDHILIRFGEISTGKNRKSFTIERLKQNTIRLVLDQYPLAKTFSNNRDMITTLINGDEFE 60

Query: 71 EVASSLKRIPIGIAQSPSPKVEKNVDLTKAVGEIMTSVYKDGMTFKITAKRSDSFEID 130
 + LK++FGIQ+FS + K + +D + + + YK G TFK+ KR+ FEID
Sbjct: 61 ALFPHLEQVFGIQSFLAIKCDRLDDIKATLKAIKDQYKPGDTFKVATKRAYKQFEID 120

Query: 131 SRALNHTLGDVAVSVLPNIKAQMOPDINKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
 + +N +G + + ++ FDI L++EIR+EA +++ D +GAGGLPVGT++GK
Sbjct: 121 TNQMAEIGGHILANTGLTVDVNPDIPLRIEIREAATPLTIRDRGAGGLPVGSAGKA 180

Query: 191 MMLSGGIDSPVAGYALKRGVDIEAVHFASPPYTPSGALKKAGHLTRKLTKEGGNIQFI 250
 MMLSGG DGFVAG+ A+KRG+ +EAVHF SPPYTS A +K DL + L++FGI++
Sbjct: 181 MMLSGGDFSPVAGFYAMKRGSLVEAVHFFSPYTPSERAKQKMDLAKLSRPGGSMTIA 240

Query: 251 EVFPETRIQREIKAKAPAYIMTLTRRPMWRITRIRDRNGLVINGESLGQVASTLES 310
 VEPT+ GE I+ + FE Y MT TR M++I DRIR ENGL II GESLGQVASTLES
Sbjct: 241 IVFPETRIQREIKQIPENTYMTATRIKMLQIADRIEKRGLAIITGESLGQVASTLES 300

Query: 311 MQAINAVTATPIIRFVVMIDKLRIIDIAQKIDTFDISIQPFEDCCTIFAPREK 364
 M AINAVT+TPI+RP+ MDK EII+ +++I T++ SIQPFEDCCTIF +P+
Sbjct: 301 MYAINAVTSTPIIRPLIAMDRTKRIEKSREIGTYETISIQPFEDCCTIFTAKPR 354
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1865> which encodes the amino acid sequence <SEQ ID 1866>. Analysis of this protein sequence reveals the following:

```
35 Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4897(Affirmative) < succ>
40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 316/404 (78%), Positives = 362/404 (89%)

Query: 11 MCYSEIMIRYGRISTKKNNRMRPINKLKNMHEVLSIYPDVSVKITDRGRHVILNGTDYH 70
 M YSEIM+R+GELSTK KNRMRPINKLKNM++ VL+ +P ++V+DRDR HV LAGTDY
Sbjct: 1 MDYSEIMVRHGLSTGKNNRMRPINKLKNNIQDVLAPEFAITVRSDRDRTHVLAGTDYQ 60

Query: 71 EVASSLKRIPIGIAQSPSPKVEKNVDLTKAVGEIMTSVYKDGMTFKITAKRSDSFEID 130
 + E+LK +PG+QA SP +K+EK+V LV AVQ+IMTS+Y+DG+TFKI KRSDH+FEID
Sbjct: 61 PIVEALKLVPGVQALSPVYKLEKSVPLLVTAVQDINTSLYRDLGTFKIAKRSDEIAFEID 120

Query: 131 SRALNHTLGDVAVSVLPNIKAQMOPDINKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
 SR LN LG AVF VLPNI+AQMK PD+ LKVEIRDEAAYISYE+I+GAGGLPVGTSGKG
Sbjct: 121 SREINSLGGAVFEVLNPIQCAQMKGPDVTLKVEIRDEAAYISYERIKSGGLPVGTSGKG 180

Query: 191 MMLSGGIDSPVAGYALKRGVDIEAVHFASPPYTPSGALKKAGHLTRKLTKEGGNIQFI 250
 MMLSGGIDSPVAGYALKRG+DIE VHFASPPYTPSGAL KA DLTR+LT+PGGNIQFI
60 Sbjct: 181 MMLSGGIDSPVAGYALKRGDLIEVHFASPPYTPSGALAKAQLTRKTRIRVGGNIQFI 240

Query: 251 EVFPETRIQREIKAKAPAYIMTLTRRPMWRITRIRDRNGLVINGESLGQVASTLES 310
```

-685-

```

 EVPFTEIGESIK KAPEAYIMILTRRFMRITD IRE R GLVI+NGESLQVASCSTLES
 5 Subjct: 241 EVPFTEIGESIKKAPAYIMILTRRFMRITDIREQRAGLVVNGESLQVASCSTLES 300
 Query: 311 MQAINAVTTPPIIRPVVTMDKLEIIDIQAQKIDTFDISIQPFEDCCCTIPADPRKTNPKK 370
 MQAINAVTTPPIIRPVVTMDKLEIIDIQAQKIDTFDISIQPFEDCCCTIPADPRKTNPKK+
 10 Subjct: 301 MQAINAVTTPPIIRPVVTMDKLEIIDIQAQKIDTFDISIQPFEDCCCTIPADPRKTNPKLG 360
 Query: 371 NTEQVEKRMVSGLVGERAVAGIMVTTIQPADSDVDVLDLIDL 414
 N E+YE+ D+GLV+RAV+GI+VT I P+ +D+V+LID IL
 10 Subjct: 361 NAEKTECFDIDGLVQRAVSGIVVTEITPEIVNDEVENLIDL 404

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 603

- 15 A DNA sequence (GBSx0643) was identified in *S. agalactiae* <SEQ ID 1867> which encodes the amino acid sequence <SEQ ID 1868>. This protein is predicted to be nifs protein homolog , fragment. Analysis of this protein sequence reveals the following:

```

 Possible site: 47
 20 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 131 - 147 (131 - 147)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1107 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

 >GP:CAA43493 GB:X61190 nifs-like gene [Lactobacillus delbrueckii]
 Identities = 177/353 (50%), Positives = 234/353 (66%), Gaps = 1/353 (0%)
 30 Query: 14 PEVLRTYQVASKIYGNPSSLHLELGTSSRILEASRKQIASLLEIAKANEIPTSOGGTEAD 73
 P+ L TY +V +KI+GNPSSLH+LG + +LEASRKQ+A LL + +EI+PTSOGGTE++
 Subjct: 3 PKALETYSQVVTIKIWNPPSLHKLGDRAHGLLEASRKQVADLLGVNTEIYPTSOGGTEBN 62
 35 Query: 74 NWVINGLAFERKQFQGNHIIVSDIEHPAVKESAKMLGEYGEIDYAPVDDKGFVDVEALVK 133
 N IKG A+ K+ FG HII S +EH +V + L GF + PVD +G V+ E L
 Subjct: 63 NTAIKGTAWAKREFQKHIIITSSVEHSAVNTFTLENLGFVRLPVDKGRKNPEDLKA 122
 40 Query: 134 LIKPETILISIMAINNEIGSIQPIKAISDLSDSKPTISFHVDAVQAIGKIPTKDYLTERV 193
 + +T L+SIM +NNEIG+IQPIK IS+L+D P I FVVD VQA+GK T RV
 Subjct: 123 ALDKDTTLVSIWNGNNEIGTIQPIKEISEILLADYNIHFHVDVQALGKIMQVTSRV 182
 45 Query: 194 DFASFSSHKPHGVRGVGFLYIKBKGRKISPLLTOGGQSTDLRSTSTENVAGIANAKALRMV 253
 D SFSSHKPHG RG+G LY K G+ + PL GGGQR LRS TEN+A IAA AKA R++
 Subjct: 183 DMSFSSHKPHGPRGTGILYKGRGRLMLPLCEGGQCEKGLRSCTENLAATAAKAKARLL 242
 50 Query: 254 MDKEVVAIFKISKMKTIHDELAKYEDITLPSG-KEDFSPNIIITFGIKGVREVLVHAFE 312
 + E + +K I LA I +PS K DP+P+I+ F ++G+RGE LVH E
 Subjct: 243 LYTEKREKADREYAIKEKISKYLAGKPGIHIPLKADFAPHLICFALGGRGTETLVHLE 302
 55 Query: 313 GHDIPISTTSACSSKAGKPGAGTLIAGISTKLACTAVRISLDDNDMQGVQVF 365
 DI+ISTTSAC+SK A TL+AM +A +NVR+S D+ N + + ++F
 Subjct: 303 QDDIYISTTSACASKKAEASTLWAKTDPDAIATSAVRISFDRSNTLRRADF 355

```

- 55 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1869> which encodes the amino acid sequence <SEQ ID 1870>. Analysis of this protein sequence reveals the following:

```

 Possible site: 41
 >>> Seems to have no N-terminal signal sequence

```

-686-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3067 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below:

Identities = 268/370 (72%), Positives = 322/370 (86%)

Query: 1 MIYFDNAGATTIPYFEVLKTYQEVASKIYGNPSSLEIGLTTSRILEASRRQIASLLELKA 60  
 MIYFDN+ATTIPY E L+TYQEV+A+KIYGNPSSLH+LGT +SRILEASRRQIA LL +K+  
 Sbjct: 1 MIYFDNATTIPYGRALKTYQEVATKIYGNPSSLHQLGTNARSILEASRRQINGLLOVKS 60

Query: 61 NEIFPTSGGTADNWNVIGLAFERQFNGNIIIVSDIEHPAVKESAKWLGEYGFIDYAPV 120  
 EIFPTSGGTE+ NW IKG+AFRK FG HII+S IHPAV ES KWL GFE+ YAPV  
 Sbjct: 61 EEIFPTSGGTESANNWAIKGLAFERQAFGKHIIISALHPAVSESVKWLLTQGEVSYAPV 120

Query: 121 DDGKGVDFEALVGLIKPETILISIMAINNEIGSIQPIKAISDLLSKPTISPHVDVAQAI 180  
 +G VDV AL +LI+P+TILISIMA+NNE+G+IQPI+AIS+LL+++PTI+PHVDVAQAI  
 Sbjct: 121 TTQGVVDVNLAEILIPDTILISIMAVNEMQAIQPIRAISNLLANQPTITPHVDVAQAI 180

Query: 181 GKIPDKYLTVERVDFSPSSHKFHVGVGVGLYIKGKRISPLTGGQQDTLRSITENV 240  
 GKIP DY+T RVD ASFS HKPH VRGVGLY K GKR++PLL+GGQQ +LRSTENV  
 Sbjct: 181 GKIPDCYMTNRVDLASFSGHKFHVGVGVGLYIKGKRINPLLSGGQQBQLRSITENV 240

Query: 241 AGIAATAKALRMVMDKEVAIPKISIKMTIHDLEACYEDITLFSKGEDPSNIIITFGIK 300  
 AGR+ AKALR+V +K+V +PK++ M+ +I+ L+ Y D+T+FS +E F+PNI+TFGI+  
 Sbjct: 241 AGIASMAKALRIIVTEKQSVLPKLTAHRDVIYKLSAYPDVTVFSAQBGFPANILITFGIR 300

Query: 301 GVRGEVLVHAFSGHDIFISTTSACSSKAGKAGTLIAMGISTKLAQTAVRISLDDDDMG 360  
 GVRGEV+VHAFE ++I+ISTTSACSSKAG+PAG+L+AMGI K AQTAVRISLDDDDMG  
 Sbjct: 301 GVRGEVIVHAFKTYKIYSTTSACSSKAGFAGSLVANGIPVKTAQTAVRISLDDDDMG 360

Query: 361 QVBQFLTIPIK 370  
 QVBQFLTIPI+  
 Sbjct: 361 QVBQFLTIPIQ 370

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 604**

40 A DNA sequence (GBSx0644) was identified in *S.agalactiae* <SEQ ID 1871> which encodes the amino acid sequence <SEQ ID 1872>. Analysis of this protein sequence reveals the following:

Possible site: 29  
 >>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1539 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 605

A DNA sequence (GBSx0645) was identified in *S. agalactiae* <SEQ ID 187> which encodes the amino acid sequence <SEQ ID 187a>. This protein is predicted to be glutathione reductase (gor). Analysis of this protein sequence reveals the following:

```

5 Possible site: 23
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.25 Transmembrane 170 - 186 (169 - 187)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.2699 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GF:BAA76640 GB:AB019579 glutathione reductase (GR) [Streptococcus mutans]
 Identities = 274/450 (60%), Positives = 346/450 (76%), Gaps = 1/450 (0%)

 Query: 1 NSKQYDIYVIGGSSAGSGTANRAAMYGAQVLLIEGGQVGGTCVNLGCVPKKIMYGAQVS 60
 N+KQYDIYVIGGSS G +ANRAAM+GAQV+L EG QVGGTCVN+GCVPKK+MYGAQV+
 Sbjct: 1 NTKQYDIYVIGGSSGSIASNRAAMHGAQVLLIEGGQVGGTCVNVGCVPKKIMYGAQVA 60

 Query: 61 ETLHKYSSGQYGFVNINFDFTILKANRDAYVQRSSQYANFERNNGVEKIDGFARFIDN 120
 ET++ Y++ YGF+V F F LK NR AY+ R + SY F+ NGVE++ +A F+d
 Sbjct: 61 ETINNYAADYGFDTITQTFHFDALKNRQAYDIRIQDSTERGFDNGVERVTSYATFVDA 120

 Query: 121 HTIEVNGQCYKAFHITTIATGGHFLYDPIIGSELGETSDDFFGNETLAFDSILIVGAGYIAA 180
 HT+EV G+ Y APHI IATGGH L PDI GSE G TSD FF + +F +VGAGYIA
 Sbjct: 121 HTVEVAGEHYTAFHILATGGHALLFDPIGSEYGITSDGFFELDAIFKRTAVVGAQYIAV 180

 Query: 181 ELAGVNVNGLGVEHIALFRKDHILRGFDPMVTSEVMAEMKSGISLHANVFKSLKDEGG 240
 E++GV++ LG ETHL R+d LR PD + ++ EM+K G LH VPK + ++
 Sbjct: 181 EISGLVHALGGETHLVRDRRPLKFKDKETVGLVDSMKKDGHLHTFSVPKVKINIDN 240

 Query: 241 KLIFEAENGKTIIVDRVWAIGRPNV-DMGLENTDIVLNDKGYIKADEFNTSDGVGYA 299
 L ENG+ VD +IWAIGR N LE T + L+ +G+I D FENT+V+G+YA
 Sbjct: 241 SLTILILEGSEYTDVLIWAIGRAANTGPNLEVTGVTLDSEGFATDAFENTVSGLYA 300

 Query: 300 IGVNNGKIALTPVAIAAGRRILSERLPHKDKNEKLOYRNVRSVIFTHFVIGTGLSEAAAI 359
 +GDVNGK+ LTPVA+ AGR+LSERLPHK K+DY +V +VIP+HPVIG+GLSE +
 Sbjct: 301 LGDVNGLKELTPVAIAAGRRILSERLPHKDKNEKLOYRNVRSVIFTHFVIGTGLSEAAAI 360

 Query: 360 EQFGRENKIVTSTFTSMYTAVTNQAQVOKLITLGEKEKVLGHGKNGXGIDENIQGSF 419
 +Q+GE+N+ VT STFTSMYTAVT++QA KML+T+G++EK++GLHG+GKG+DMIQGSF+
 Sbjct: 361 DQNGEENVIVYRSTFTSMYTAVTSRQACKMKLVTVGRDKIVLGHGIGYGVDENIQGFA 420

 Query: 420 VAIKMGATKADFDVVAIHPTGSEEFVIMR 449
 VAIKMGATKADFD+VAIHPTGSEEFVIMR
 Sbjct: 421 VAIKMGATKADFDVVAIHPTGSEEFVIMR 450

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 187b> which encodes the amino acid sequence <SEQ ID 187c>. Analysis of this protein sequence reveals the following:

```

50 Possible site: 23
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.33 Transmembrane 173 - 189 (173 - 191)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.1532 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

60 An alignment of the GAS and GBS proteins is shown below:

-688-

Identities = 268/446 (60%), Positives = 340/446 (76%), Gaps = 1/446 (0%)

Query: 5 YDYIVIGGGSGGTANRAAMYGARVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVETLH 64  
 YDYIVIGGGSGG +ANRAAM+GARVLL EG ++GGTCVNLGCVPKK+MWYGAQV++ L  
 5 Sbjct: 8 YDYIVIGGGSGAGIAGANRAAMHGAARVLLABGKIEIGTTCVNLGCV+KKVMWYGAQVADILG 67

Query: 65 KYSSGYGFVNNLNFDTTLKANRDAYVQRSGQYANPERNGVEKIDGPARFINHTIE 124  
 Y+ YGF+ FDF LKANR AY+ R SY FE+NGV++I +A F D HT+E  
 10 Sbjct: 68 TYAKQYGFDFKEKAFDFKQKANRQAYIDRIHASYERGFEQNGVERIDYAVFQDAHTVE 127

Query: 125 VNGQQYKAPHITATGGHPLYPDILGSELGETSDDFPGWETLPDSILIVGAGYIAELAG 184  
 + GQ Y APHI IATGHPH++PDI G++ G+SD FF ++P +VGAGYIA ELAG  
 10 Sbjct: 128 IAGQLYKAPHILATGHPHFPDIEGAQYGLISDGFALDEVFKRTAVVGAGYIAVELAG 187

Query: 185 VVNELGVETHLAFKRDHILAGFDNVTSEVMAMEKSGISLHANEVPSKLRDSQKLI 244  
 V++ LG +T L R D LR FD + ++ EM +G LH + ++ L  
 15 Sbjct: 188 VLHALGSKTDLFIRHRRPLRSFRTIVDVLVDENAVNGPRLHTHAERVAKVNTIDESLTL 247

Query: 245 EAENGKILVDVRVMAIGRGPVVD-MGLENTDIVLNDKGYIKADEFENTSDVGVAIGDV 303  
 ++G+ VD++IWAIGR PH++ L+ T +LNDKGYI+ D +ENTSV G+VAIGDV  
 20 Sbjct: 246 YLKDGQSEVVDQLIWAIGRPHLEGSFLDKGYIVTINDKGYIETDAYENTSVIGYIANGDV 307

Query: 304 NGKIALTPVAIAAGRLSERLSEKDKNEKLDYHNVPSVITPHVIGTVGLSEAAIEQFG 363  
 NGK+ALTTPVA+ANGRLSERLPN K +EKLDY NV +VIF+HPVIG+VGLSE AA++Q+G  
 25 Sbjct: 306 NGKIALTPVAIAAGRLSERLSEKDKNEKLDYQNVATVIFSHFVIGSGLSEAAVCKY 367

Query: 364 EDNIKVYTSPTSMYTAVTINRQAVOMKLTIGKEEKVIGLAGVGYGIDEMIQSFVAIK 423  
 ++ +K Y S PSM+TA+T +RQ MKL+T+G EK++GLAG+GYG+DEMIQSF+VAIK  
 30 Sbjct: 366 QEAVKTYQSRPTSMPTAITNRQPCMLKLVTVGTERKIVGLHIGYGVDEMIQSFVAIAIK 427

Query: 424 MGATKADFDFTVAIHPTGSEEFVIMR 449  
 MGATKADFD+TVAIHPTGSEEFVIMR  
 Sbjct: 428 MGATKADFDNTVAIHPTGSEEFVIMR 453

35 SEQ ID 1874 (GBS417) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 5; MW 53kDa).

GBS417-His was purified as shown in Figure 216, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 606

A DNA sequence (GBSx0646) was identified in *S.agalactiae* <SEQ ID 1877> which encodes the amino acid sequence <SEQ ID 1878>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3122 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AA62417 GB:AF084104 hypothetical protein [Bacillus firmus]  
 Identities = 33/110 (30%), Positives = 66/110 (60%)

55 Query: 1 NANNVYDLANELERAVRALPEYQVLTAKSAIESDADAQVLAQDFLATQSKVOEMQSSQM 60  
 N+NVYD A+EL++A+ E+ A+ + IE+D A+ ++F Q ++Q+ G  
 Sbjct: 1 MSNVYDKAHELKCAIARSEFSAKSMHREIRADEIAKMKLENFIRNLQLELOQKMQGIG 60

Query: 61 PSQERQDEMSEKIGKIEKSNLLKVYFDQOQRLSVYMSDIEKIVFAFMQDL 110

-689-

++EE + + E ++ ++L+ + +QRLSV + DI KI+ P+++  
 Subjct: 61 ITREARQKQQQFELVQKHELISKLMRAEQRLSVIGDINKLITPLEEI 110

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1879> which encodes the amino acid sequence <SEQ ID 1880>. Analysis of this protein sequence reveals the following:

Possible site: 38  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4058 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/108 (62%), Positives = 86/108 (78%)  
 Query: 4 VYDTANLEAEAVPALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQPMQSGQHPQ 63  
 +YD AN+LEAEAVPALPEYQ VL K AI++D A L+ +F+A Q K+Q NMQSGQMP+  
 Subjct: 5 IYDYARQLERAVRALPEYQKLEVKRAIQADVASSELDFEYFVAMQKEIKQMMQSGQNPTA 64  
 Query: 64 EEDBMSKLGKETIESNDLLKLVFDQQRLSVYMSDIEKIVFAPMQLM 111  
 EEQ + +L +KIE+ND LK YF+ QQ LSVYMSDIE+IVFAP++DL+  
 Subjct: 65 EEQTSIQELSKETIANDQLKAFEMQQLSVYMSDIERIVFAPKDLV 112

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 607

A DNA sequence (GBSx0647) was identified in *S.agalactiae* <SEQ ID 1881> which encodes the amino acid sequence <SEQ ID 1882>. This protein is predicted to be chorismate synthase (aroC). Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.67 Transmembrane 343 - 359 ( 341 - 364)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2869 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]  
 Identities = 227/381 (59%), Positives = 282/381 (73%), Gaps = 2/381 (0%)  
 Query: 1 MRYLTAGESHGPSLTAIEGIPAGLKLSAKDINELAKRRQGGYGRNRMKLIETDQVISS 60  
 MRYLTAGESHGP LT IIEG PA L+L A DIN DL RRQGG+GRG RM+IE DQV I  
 Subjct: 1 MRYLTAGESHGPQATTIEGAPAEQLLEVAQDINVDIARRQGGHGRGGRMIEKDQVIG 60  
 Query: 61 GVRHGKTLGSPITLIVTNKHSKNLIDMSVEDI--KERLQKRRIKHPRGHADLVGGIK 118  
 G+REKKT G+PI L V NKD W IM E + +E + KR+I PRPHADL G I K  
 Subjct: 61 GIRGKKTGAPALAVVENKDKHWTKLMGASPLTGDEREKIRKRIKTRPRGHADLVNGAIK 120  
 Query: 119 YRFDDLKALERSSARETTIMRVAIGAIKRLKIEGIEIANHIVFGGKEITVFDKLTQV 178  
 Y D+RN LERSSARETT+RVA GA+AK+IL+ GIE+ +H++ GG + +  
 Subjct: 121 YGRDRMNVLERSSARETTVRVAAGAVAKKILKLETFGLVGVSHLEIGGVKAEKTSYDQLS 180  
 Query: 179 QIKVLSSQSQVAIVNPSFQEIKDVIDSVKKAGDTIGGVVETIVGGVPVLGSGYVHMDRK 238  
 +K L+ S V ++ EQE+ ID K GD+IGGVVE IV GVP+GLGS+VH+DRK  
 Subjct: 181 NLKELAEASPVCLDKAEQEMLAIDQAKGNGDSIGGVVEIVGVGVPVGLGSHVHYDRK 240

-690-

Query: 239 L DAKLAQAVVSNINAFKGVFPGFGKSGFLKSGVMSISWTQGGYIRGSNNLGGFEGGM 298  
 LDAKIA AV+SINAFKGVFPG+GF++ GS+V D I+W +++GY R+SNNLGGFEGGM  
 5 Sbjct: 241 L DAKIAA VSNINAFKGVFPGFGFPAASKPGSEVHDEIANDBERGYRKSNLGGFEGGM 300

Query: 299 TNGSEPIVRGVGMKPIPTLYKPLMSVDIDTSHEPYRATVERSDPTALPAAGVMEAVVATVL 358  
 TNG PI+VRGVGMKPIPTLYKPL SVDI T EP+ A++ERSD A+PAA VV RAIVA  
 Sbjct: 301 TNGSEPIVRGVGMKPIPTLYKPLQSVDIATKRPFAASIERSDSCAVPAAVAEVAVAVEV 360

10 Query: 359 VTEVLEKFPSSINNYELKSAVK 379  
 +LE+ F +D + E+++ ++  
 Sbjct: 361 ANALLERFGADQVEIEKNIR 381

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1883> which encodes the amino acid  
 15 sequence <SEQ ID 1884>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.75 Transmembrane 342 - 358 ( 342 - 359)  
 20 INTEGRAL Likelihood = -0.16 Transmembrane 155 - 171 ( 155 - 171)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1298 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA05375 GB:AP001512 chorismate synthase [Bacillus halodurans]  
 Identities = 213/390 (54%), Positives = 277/390 (70%), Gaps = 2/390 (0%)

30 Query: 1 LRYLTAGESHGPSLTALIEGIPAGLTLPADINHLQRQGGYGRGARMISITDRVQISS 60  
 +RYLTAGESHGP LT IIEG PA L L DI+ +L RRQGGYGRG RM IE D+VQI  
 Sbjct: 1 MRYLTAGESHGPSLTALIEGAPALEVADINVDLARRQGGHGRGRMIQKDQVQIVG 60

35 Query: 61 GVRHGKTTGAPITLTVINKDHQKNLQVMAVGDI--ETTLKLRVRVHPRPHADLGVGIIK 118  
 G+RHGKTTGAPI L V NKD + W +M + +E ++KR++ PRPHADL G IK  
 Sbjct: 61 GIRHGKTTGAPIALVVENKDKHHTWKIMGAEPLTQDEKEIKRKITRPRPHADLGAIIK 120

Query: 119 YHFNDELALERSARETTMVAAGAVAKRILAEGLIDMLHHILIFGGITITITPSKLSFR 178  
 Y D+R+ LERSARETT+KVA GAVAK+IL GI++ H+L GG+ S  
 40 Sbjct: 121 YGHRDMENVLERSARETTVRAAGAVAKKILRTFGIEVGSVHLKIGGVIAKTSYDQLS 180

Query: 179 ELQERALASHELSTVNPQKSEIKTYIDIKIKEGDITGGIETITVCGVPAGLSGYVOMDKK 238  
 L+E A S + ++ + E+E+ ID+ K+ GD+IGG+E IV+GVP GLGS+V +D+K  
 45 Sbjct: 181 NLKELAEASPVRCLOKEAGQMTAIDQAKENKDSIGGVVEVIEGVPIGLSHVHYDERK 240

Query: 239 L DAKLAQAVVSNINAFKGVFPGFGKSGVMSISWTQGGYIRGSNNLGGFEGGM 298  
 LDAKA AV+SINAFKGVFPG GF+ + GS+V DEI W +GY R++N+LGGFEGGM  
 Sbjct: 241 L DAKIAA VSNINAFKGVFPGFGFPAASKPGSEVHDEIANDBERGYRKSNLGGFEGGM 300

50 Query: 299 T TQGPLVVGVMKPIPTLYKPLMSVDIDTSHEPYKATVERSDPTALPAAGVIMENVATVL 358  
 T G F+V+V+GMKPIPTLYKPL SVDI + EP+ A++ERSD A+PAA V+ E VVA +  
 Sbjct: 301 TNGSEPIVRGVGMKPIPTLYKPLQSVDIATKRPFAASIERSDSCAVPAAVAEVAVAVEV 360

Query: 359 AKELLETFSSTIMSELQKAPSDYRAYVQF 388  
 55 A +LE F + + E++K ++ + F  
 Sbjct: 361 ANALLERFGADQVEIEKNIRPNKARLF 390

An alignment of the GAS and GBS proteins is shown below:

Identities = 284/388 (73%), Positives = 333/388 (85%)

60 Query: 1 MRYLTAGESHGPSLTALIEGIPAGLTLPADINHLQRQGGYGRGARMISITDRVQISS 60  
 +RYLTAGESHGPSLTALIEGIPAGL L DI+ +L+RRQGGYGRG RM IKTD+V ISS  
 Sbjct: 1 LRYLTAGESHGPSLTALIEGIPAGLTLPADINHLQRQGGYGRGARMISITDRVQISS 60



-691-

Query: 61 GVRHGKTLGSPITLTVINKDHSKWLIDMSVEDIKERIKQKKRIKHPKQHADLVGGIKYR 120  
 GVRHGKT G+PITLTV NKDH KWLID+M+V DIES LK KKR+KHPKQHADLVGGIKY 120  
 Sbjct: 61 GVRHGKTLGAPITLTVINKDQKWLIDVMAVDIEETLKLKRRVKKHPKQHADLVGGIKYH 120

5 Query: 121 FDDLRLALERSSARETINRVAIGAIAKRILKEIGETIANHIVVFGGKEITVPDKLVQOI 180  
 F+DLR+ALERSSARETINRVA+GA+AKRIL E+GI++ +HI++FGG IT+P KL+ +++  
 Sbjct: 121 FNDLRDALERSSARETINRVAVGAIAKRILAEIGDIDLAHILIFGGITITIPSKLSFREL 180

10 Query: 181 KVLSSQSQVAIVNHSPEQEIKYIDSVKKAGDTIGGVVETIVGVVPVLGSSVHMDKLD 240  
 + + S+++IVNP E+EIK YID +KK GDTIGG++ETIV GVP GLGSTV ND+KLD 240  
 Sbjct: 181 QERALHSELISVNPQEEETIKYIDKIKKBDTIGGIIETIVQGVPAAGLSTVQVMDKLD 240

Query: 241 AKIAQAVVSINAFKGVFGLGFKSGFLKGSQVMDISWTKDQGYIRQSNLQSGFBGSMTN 300  
 AK+AQAV+SINAFKGVFVG GF GF KGSQVMD I+WI QGY RQ+N+LQGFGBGSM 300  
 Sbjct: 241 AKLAQAVLSINAFKGVFVGAGFGMGFGKGSQVMDIETIWTPTQYGRQTNHLOGFBGSM 300

15 Query: 301 GEPIIVGVGMKFIPTLYKPLMSVDIDTHEPYRATVERSPTALPAAGVVMHVAIVATVLT 360  
 G+P++V+GVMPKFIPTLYKPLMSVDID+HEPT+ATVERSPTALPAAGV+ME VVATVL 360  
 Sbjct: 301 GQPLVGVGMKFIPTLYKPLMSVDIDSHPTKATVERSPTALPAAGVIMHVVATVIAK 360

20 Query: 361 EVLEKFGSSNMYELKEAVKLYRYVDFP 388  
 E+LE FSS M EL+AA YR YV F  
 Sbjct: 361 EILETFSSSTMSLQKAFSDYRAIVKQF 388

25 A related GBS gene <SEQ ID 8617> and protein <SEQ ID 8618> were also identified. Analysis of this protein sequence reveals the following:

Lidop: Possible site: -1 Crend: 9  
 MoG: Discrim Score: -2.42  
 OvH: Signal Score (-7.5): -3.23  
 30 Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -4.67 threshold: 0.0  
 INTEGRAL Likelihood = -4.67 Transmembrane 343 - 359 ( 341 - 364)  
 PERIPHERAL Likelihood = 0.69 214  
 35 modified ALOM score: 1.43  
 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 40 bacterial membrane --- Certainty=0.2869(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 57.7/73.8% over 354aa Bacillus subtilis  
 EGAD[20299] chorismate synthase Insert characterized  
 SP|P31104|AROC\_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE  
 PHOSPHOLYASE)  
 50 (VEGETATIVE PROTEIN 216) (VEG216). Edit characterized  
 GP|143806|gb|AA20859.1||M80245 AroF Insert characterized  
 GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase Insert characterized  
 PIR|C69590|C69590 chorismate synthase aroF - Insert characterized

55 ORF00121(301 - 1359 of 1719)  
 EGAD[20299]BS2267(1 - 355 of 366) chorismate synthase {Bacillus  
 subtilis}SP|P31104|AROC\_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-  
 PHOSPHATE PHOSPHOLYASE) (VEGETATIVE PROTEIN 216) (VEG216).GP|143806|gb|AA20859.1||M80245  
 AroF {Bacillus subtilis}GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase {Bacillus  
 60 subtilis}PIR|C69590|C69590 chorismate synthase aroF - Bacillus subtilis  
 %Match = 35.0  
 %Identity = 57.6 %Similarity = 73.7  
 Matches = 204 Mismatches = 92 Conservative Sub.s = 57

65 75 105 135 165 195 225 255 285

-692-

```

IQLSRVAERKNLMRPGISQDIYNNCLKPLFVHYAENKDLVFDILSHDKKASGQFIKIVILPQLGSATVHPILEERD
315 345 375 405 435 465 495 525
YLER*MRVLTAGESHGSPSLTALTEGIPAGLKLAKNDINEKLRRGGVYGRGRWKIETDQVITISSGVHKGTLGSDITLT
5 ||||| ||||| ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 MRYLTAGESHGSPQLTTITISVFPAGLYITTEEDINPELARAQKEHGRGRMRQIEKDDQAKIMSGVVRHARTLGSFTALV
 10 20 30 40 50 60 70

555 609 639 669 699 729 759
VTNRKHSKWLDMISVEDI--EERLKQKRRIKHPRFOHADLVGQIKYRFPDILRNALERSASRETTMRVAIGAIKRLIKEI
10 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 VENNDWKHWTKIMGAAPITDEEEKEMKQISRPFGHADLNGAIKYNHIDMRNVLERSSARETTVRVANGAVAKILSEL
 90 100 110 120 130 140 150

769 819 849 879 909 939 969 999
GELIANHVTVPFGKEITVFDKLTVOQIKVLSSQSAIVNPSFEGEIKDYIDSVKKAGDTIGAVVETIVGVFVGLGSYV
15 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 GIKVAGHVLQIGAFKAKTGVTSLDLQVTEESPVRCYDEAGKUMAAIDENKANGDSIGGIVETIVGVFVGVGLGSYV
 170 180 190 200 210 220 230

1029 1059 1089 1119 1149 1179 1209 1239
HMDRKLDAKLAQAVVSIINAFKGVPEGLFKSGPLKGSQVMSISNTKDGQYIROSNLGGFGGVINGEPTIVRGVMKPTI
20 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 HYDRKLDSEKLAALVSLINAFKGVPEGLFGEAAGRNKSGVHDEIINDDEKGYTRATNLSLGSKGVTTQMPITVIRGVKPTI
 250 260 270 280 290 300 310

1269 1299 1329 1359 1389 1419 1449 1479
FTLYKPLMSVDIDTHEFYRACVVERSDPTALPAAGVMEAVVATVLTVELEKFPSSDNMN*KKL*NYIAMIIFR*KLV
30 ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 FTLYKPLKSGVDIEKFPFASIERSDSCAVPAASVVAALSLGKLQPSLNSD
 330 340 350 360

```

SEQ ID 8618 (GBS192) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 4; MW 44kDa).

GBS192-His was purified as shown in Figure 196, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 608

A DNA sequence (GBSx0648) was identified in *S.agalactiae* <SEQ ID 1885> which encodes the amino acid sequence <SEQ ID 1886>. This protein is predicted to be 3-dehydroquinate synthase (aroB). Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.82 Transmembrane 99 - 115 (98 - 116)
45 ----- Final Results -----
 bacterial membrane --- Certainty=0.2529 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
Identities = 138/351 (39%), Positives = 200/351 (56%), Gaps = 4/351 (1%)
55 Query: 3 VEVDLNPHPYHIKISEGCFSEAGDVSILWQKQMITIITDSNVEILYAGESLWNLKQKGF 62
 + V LP PY ++I G + D ++ L + I +++ + YGE ++ L++ G+
Sbjct: 5 IPVPLQSPFYQVQIVFGGLAAIADHLAPLGLGKKIMVVSNEPIYDYGVEVQIALQRAY 64

```

-693-

Query: 63 TVIVPSFAAGSEAKTLEVAARIYAFIAKHIMTRSDGIIALGGVVDGLAAPVASTYVSGI 122  
V AGR KTL N +Y + + + R+ + + + LGGV + GD + F A + T + RGI  
Sbjct: 65 EVFQHLIPAGETHKTLASINELNDVAFQANLERNSTLSLGGGVIGDMTGFGAATWLRGI 124

5 Query: 123 HFLQIPTSLSIAQVDSIGGKTGWTSTFAKNVGTFAQPGVLDIPVTKLGNRLVZGM 182  
+ F + Q + PTL A VD + SIGGKTGM KN + G F Q P V IDPV LKTL RE GM  
Sbjct: 125 NFVQVPTSLAMVDASIGGKTGVNHPQGNLIGAFYQRLVYIDPVVTKLTPERFRAGM 184

10 Query: 183 GEVIKYGLIDDIKMLHILEMD - GTDISILONALA - IYHSCOVKRKHVLADQYDEGLR 239  
EVIKYGI D + L+ LEE + + ID + D L II SCQ K V D + + GLR  
Sbjct: 185 AEVIKYGVINDSELFTALEEAEDLSSIDRLPDELLTKIIQRSCQAKVDVVSDQENKAGLR 244

Query: 240 MHLNPGHTIGHAIEVHMGYGEIMHGEBAVAIGMQLSRVAERKNLAPRGISDIYNCLAF 299  
LN + GHT + GH + E GYG I HGEBAVIGM + + + A L + + + + LK  
Sbjct: 245 AILNYGHTVGHGVSLSITGYGVNHSGEBAVIGMEAAAKIAHYLGCDQSIDGRQQLLKT 304

15 Query: 300 GLP/PHY - AKMDKVLFDLISHDKKASQOFIKVILPOLGSATVHOIPEEM 349  
LP + L L HDKK + + + + G T + + E +  
Sbjct: 305 KLPTMPPTLAVENLLASLLHDKKVKAGKVRFLPTAIGQVITSDAVTDEV 355

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1887> which encodes the amino acid sequence <SEQ ID 1888>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence  
25 INTEGRAL Likelihood = -0.43 Transmembrane 97 - 113 ( 97 - 114)  
----- Final Results -----  
bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAAL8068 GB:D90911 3-dehydroquinate synthase [Synecocystis sp.]  
Identities = 123/349 (35%), Positives = 190/349 (54%), Gaps = 9/349 (2%)

35 Query: 1 MPQTLNVHSRVKDYDILPTDHLVKTADCLGERKQ - RKLFTDQTVVHYLYQTLEFEFAQ 59  
M T + V Y + L + AD L + K + + + + Y Y + + + Q  
Sbjct: 1 MATTIPVLPQSPVQVQIVPGLAAIADHLAPLGLGKKINVSNSFEIYDYGEVVLQALQ 60

40 Query: 60 Q - - YNAFVHVCPGPGQSKSLERSAIYDQLIAENFSKKMIVTIGGGVVDGLGSFVAATY 117  
+ Y F H + P G K + L + + YD N + + + + GGGV + GD + GF AAT +  
Sbjct: 61 RAGVEVFPQLIPAGETHKTLASINELNDVAFQANLERNSTLSLGGGVIGDMTGFGAATW 120

45 Query: 118 YRGIPYIQTPTLLSQVDSIGGKGVHFKGLNIMGSIYPPREALIISTFLETLPOREF 177  
RGI + + Q + P + LL + VD + SIGGK GV + N + IG + Y P + I L + TL + P + REF  
Sbjct: 121 LRGINFVQVPTSLAMVDASIGGKTGVNHPQGNLIGAFYQRLVYIDPVVTKLTPEREF 180

Query: 178 SCGISSEMLKIGFTHDRPLFPQQLRDPQ - - - - KETDKQGLERLIYQISINKKRIVEQDEF 232  
G + + E + + K G ID LF L + + + L + + I + S K + V QDE E  
50 Sbjct: 181 RAGMAEVIKYGVINDSELFTALEEAEDLSSIDRLPDELLTKIIQRSCQAKVDVVSDQENK 240

Query: 233 NGLRMSLNPGHTIGHAIESLCHDFTHHGEAIAIGMVDAKLAHSVGLLPKEDLSLLQV 292  
GLR LN + GHT + GH + ESL + + HGEA + AIGM AK + A GL + D Q +  
Sbjct: 241 AGLRAILNYGHTVGHGVSLSITGYGVNHSGEBAVIGMEAAAKIAHYLGCDQSIDGRQQL 300

55 Query: 293 PERYQLPTTLERADVSATSLFDVFKTKDNSECHIIPIFLTPTGTFTLA 341  
+ + LPT + + + + L DEK + FILPT G T + +  
Sbjct: 301 LLKTKLPTMP - PTLAVENLLASLLHDKKVKAGKVRFLPTAIGQVITSDAVTDEV 348

60

An alignment of the GAS and GBS proteins is shown below:

Identities = 121/332 (36%), Positives = 182/332 (54%), Gaps = 7/332 (2%)

Query: 12 YHIKIEGCFSEAGDWSHLAQKMITIITDSNVBLIYGSILVNQLKKQGFVHVSFAA 71  
Y I + + D + Q + + + ITD V LY + + L + + Q + + V

-694-

Sbjct: 14 YDILFTDHLVKTLADCLGERKQKLL- FITDQIVYHLY-QTLFEEFAQQ-YNAFVHVCP 70

Query: 72 GEASKTLFVANRYAFLAKHHMTRSDGILALGGVGVGLAASFVASTYMRGIFLQIPTSL 131  
 5 G SK+LE + IY L + ++ D I+ +GGVGVGL FVA+TY RGI ++QIPT+L  
 Sbjct: 71 GGQSKSLERVSALYDQLAENFSKKOMVTIGSGVGVGLGFGVAATYTRGIPYIQTPTL 130

Query: 132 TAQVDSBIGGKGVNITSFAKMMVCTFAQPDVGLIDPVTIKTLGNRLVSGMGVIVKGLI 191  
 +QVDSBIGK GV+ NM+G+ P+ ++I L+TL RE G+ E+K G I  
 10 Sbjct: 131 LSCVDSBIGGVGVHFGKLNMMIGSIYPPEALISTFTLTLPQREFSGSISMLKIGFI 190

Query: 192 DDIKLWHILEMDGTIDSILDNALAIYHSCQVKKHVLADQYDKGLRMHLPFGHTIGHA 251  
 D L+ L+ + D +IY S K++ V D+++ GLEM LNFQHT+GHA  
 Sbjct: 191 HDRPLFQQLRDPQKETDK--QSLERLIYQSISMKRIVEQDEFENGRLMSLNFQHTLGH 248

Query: 252 LEVHNGYGEHNGEAVAIGNIQLSRVNERNLAPRGISQDIYNMCLKFLPL--VHYAEND 309  
 IE + HGEA+HGM+ +++A K L+P+ + + ++ LP + R+  
 15 Sbjct: 249 IESLCHDFYHNGEATAIGMVDAKLAIVSGLLFKEDLSLLQVFRYQLPTTLERADVS 308

Query: 310 EDVLFDTLSHDKKASQPTIKVILPQLGSATV 341  
 LPA- DKK S Q I ++ + G T+  
 20 Sbjct: 309 ATSLDFVFKTDKQSGSHLIFLPTETOPTL 340

SEQ ID 1886 (GBS336) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 2; MW 42.7kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 5; MW 68kDa).

The GBS336-GST fusion product was purified (Figure 209, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 310), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 609

A DNA sequence (GBSx0649) was identified in *S. agalactiae* <SEQ ID 1889> which encodes the amino acid sequence <SEQ ID 1890>. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

          bacterial cytoplasm --- Certainty=0.3884 (Affirmative) < succ>  
 40           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9973> which encodes amino acid sequence <SEQ ID 9974> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA514240 GB:299116 3-dehydroquinate dehydratase [Bacillus subtilis]  
 Identities = 70/233 (30%), Positives = 127/233 (54%), Gaps = 12/233 (5%)

Query: 2 KIVVPMPSRLEE-QEIDLSKFSVDIEMRADALPK---DDIINVAPAIPKFKAGE 56  
 KI++P+M ++ ++ E + K + DI+EMR D K + + + + +

50 Sbjct: 17 KIIIPLMGKTKQLNEAAVKLLNPDIVEMRVDPFKANDREAVTKLISKIRKSLKDL 76

Query: 57 LIPTLTRECGNIVLSDAEYVELIKQINSYNDPYIDPEYPSHKVSVQEMLEFPN--- 112  
 +FT RT +RGG++ ++ Y+ L++ + D ID R PS ++

55 Sbjct: 77 FLPTFRTHKGGSGMDESSYLALLESATQTKDIDLIDRLPSGDANVKALVSLAENNV 136

-695-

Query: 113 -LVLSYHNPQETP--ENIMKIFSELTAAPRVVKIAMPKQEQVLDVMNYTRGFKTIINP 169  
 +V+8 H+P++TP + I+ ++ L + K+AMP + D+L +++ T KTI  
 Sbjct: 137 YVVMGNHDPFKETPVKDELISRIRKQQLGSAHPKMAVMPNLTGLLTLTLDATYTKKIYA 196

5 Query: 170 DQVATVSMKIGIRISRFAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALL 222  
 D+ T+SM+ G ISR +G+V GS+ TF + + APGQI +SE++ V +L  
 Sbjct: 197 DRPIITMSMARATGLISRSGSVFSGACTFGAGGEEASAPGQIPVSELRSLDIL 249

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1891> which encodes the amino acid sequence <SEQ ID 1892>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3248(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 160/225 (71%), Positives = 198/225 (87%)

Query: 1 NKIVVFMPSRLSEAEIDLSKFDSDVIEWRADALPKDDIINVAIPALFEKFAHELIPT 60  
 N+IV FVMPS +EAQ ID+SK++ V++IEWRAD LFKD+I+ VAPALFEKFAG EIIPT  
 Sbjct: 1 NRIVAFVMPRHFDGACALDISKYEDVLEIWRADFLPKDSIVAVAPALFEKFAHELIPT 60

25 Query: 61 LRTTREGGNIVLSDAEYVSLIQKINSIYNPDYIDFEYFHSKVEVPOEMLFEPNLVLSYHNF 120  
 LRT +EGGNI LS EYV++I++IN+IYNPDYIDFEYF+HK VPOEML+FPNL+LSYHNF  
 Sbjct: 61 LRTVREGGNITLSSQFYVDIKINAIYNPDYIDFEYFTHKSVFQEMLDFFNLISYHNF 120

30 Query: 121 QETPENIMKIFSELTAAPRVVKIAMPKQEQVLDVMNYTRGFKTIINPQOVATVSMK 180  
 +ETPEN+ME FSE+T LAAPRVVKIAMP+EQVLD+MNYTRGFKT+NP+Q +AT+SM K  
 Sbjct: 121 EETPENLMEAPSEMTKLAAPRVVKIAMPQEQVLDLDMNYTRGFKTIINPQETATISMGK 180

35 Query: 181 IGRISRFAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALLDAD 225  
 +GR+SRFAGDV GSWT+ LD PQQ+T++NKR+ +L+ D  
 Sbjct: 181 LGRLSRFAGDVIGSSWTVSLDHVSGPQVTLDMKRIIEVLEMD 225

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 610

A DNA sequence (GBSx0650) was identified in *S.agalactiae* <SEQ ID 1893> which encodes the amino acid sequence <SEQ ID 1894>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1195(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-696-

**Example 611**

A DNA sequence (GBSx0651) was identified in *S. agalactiae* <SEQ ID 1895> which encodes the amino acid sequence <SEQ ID 1896>. Analysis of this protein sequence reveals the following:

Possible site: 41  
>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3431 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15862 GB:Z99123 alternate gene name: ipa-19d-similar to  
hypothetical proteins [Bacillus subtilis]  
Identities = 161/396 (40%), Positives = 235/396 (58%), Gaps = 11/396 (2%)

Query: 1 MNKLLKVN SVVERKIKSGAQLEKKDFDTSLVNO----LVQLFSGSN-QFLGMAYLSPQNK 55  
M L + KIK G L+EK+ S + LV + S+S +PL Y QNK  
Sbjct: 1 MKLLTLKFAHAAKIKKGYPILEKALAGSAGHMKEGDLVDIVSSESGEFLARGYYQLQNK 60

Query: 56 GIGWLLSQQVDF-PNHDFVSLFEKSEKKQKFEKSSCTTAYKLFNQDGNFAGGLTIDFY 114  
G+GW L+R + + +F+S K+ + R K ++ TTA+RLFN +GD GG+TID+Y  
Sbjct: 61 GVGWTLTRKHQKIDQAFLSKLTKAAQRAKLEAQDTTAPRLFNSEGGVGSGVITIDY 120

Query: 115 SDYALFSWINEFVITNKMIVAAFKQVYVNIKSGAYEKIRFGKLDP---ESAHLYGQRAPE 171  
Y L WY++ +YT + M+++A ++ + K YEK RF + + G+  
Sbjct: 121 DGVLLIQWYSGKGYITFKIMLISALDRMDLDYKALVEKKRFDTAQGVVDEDDFVKGRRGEF 180

Query: 172 SFLLLENNIKYSVFINDGLMTGIFLDQHVVRKALATNLSEKKVIVNMFYSYTAAPSVAAR 231  
+I EN I+Y+V LA+G MTGIFLDQ VRKA+ ++GK VLN PSYT APSVAAA+  
Sbjct: 181 PIIIQENGIQYAVDLNKGAMTGFILQDQHVVRKAIRDYAKGKTULNTFSYTGAPSVAAAL 240

Query: 232 GGALETTSVDLAKRSRELSKARFDANQIVTINHRFIVMDVFEYKIKAKRKLISYDVID 291  
GGA +TTSVD+A RS + F N++ + H VMDVF Y+ YA +K L +D+I++D  
Sbjct: 241 GGAETTSVDVANSRLAKTIEQFSVNKLQYSAHDIKVMDVFNYSYAAKDLRFDLIILD 300

Query: 292 PPSFARNKKQTFSTVTKDYKLEQALDILTPGGTIIASTNAANLTVSQFKKQLEKGFKA 351  
PPSFAR KK+TFS KDY L+++ + I G I+ASTN+ + +FK ++ F +  
Sbjct: 301 PPSFARTKKRTFSAADYKMLKETIATDAKGVIVASTNSASGMMKFKGFDIAAPKFT 360

Query: 352 SHNYISLQ--LPEDPTINDKQQSNYLKVPFIKVK 385  
+ Y +++ LPEDF + NYLKV ++ K  
Sbjct: 361 NERYTIEREFLPEDFKTISAPFEGNYLKVVLQKK 396

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1897> which encodes the amino acid sequence <SEQ ID 1898>. Analysis of this protein sequence reveals the following:

Possible site: 29  
>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2699 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 259/386 (67%), Positives = 315/386 (81%), Gaps = 1/386 (0%)

Query: 1 MNKLLKVN SVVERKIKSGAQLEKKDFDT-SLVNQLVQLFSGSNQFLGMAYLSPQNKIGW 59  
MNKL ++S VE+K+ +G QLL++KDF NQLVQL ++SN+ +G AY+S QNKIGW  
Sbjct: 1 MNKLYIDSFVEKKLITAGVQLLEKDFSKIKKKNQLVQLVTKSRNPIGTAYISKQNKIGW 60

-697-

Query: 60 LLSRQVDFDNHIDYFVSLFPEKSRERKQKFEKSSQITAYRLFNQDGNFQGLTIDFVSDYAL 119  
 L + D + YFVSLF ++ KRF + S +T AYRLFNQ+GD FQGLTID Y D+A+  
 Sbjct: 61 YLGPKEIDLSISYFVSLFSAKAKRQDFQSGDENTAYRLFNQSGDGFQGVITDLYKDFAV 120

5 Query: 120 FSWYNFVYTHROMIVAAPKQVYENIKGAYEKIRFKGLDFESAELVQGEAPSPFLILENN 179  
 FSWYN FVY ++MI+ AFQV+P +KGAYEK RFKQ D R+NLHYG A E+P ILEN  
 Sbjct: 121 FSWYNAFVIDKKEMIMEAFQCVFFFEVKGAYEKRFKQGPOTETAHLGELAQETFSILENG 180

10 Query: 180 IKYVSFIADGLMTGIFLDQRDVRKALATNLSEKKVINMFSTYTAAPSVAAAVQGALETTS 239  
 I Y VFLN+GLMTGIFLDQRDVR+AL L+ GK +LN+FSYTAAPSVAAA+GGA+ETTS  
 Sbjct: 181 IAYQVFLKGLMTGIFLDQRDVRRAIVDGLAMGSKLNLFSYTAAPSVAAAMGGALETTS 240

Query: 240 VDLAKSRRLSKANFDANQIVTDNRHFVMDVFEYKYAKRKLSDVIVIDPPSPFARNK 299  
 VDLAKSRRLS AHP+ NQ+ +H F+VMDVFEY+KYAKRKL +DVIIVIDPPSPFARNK  
 Sbjct: 241 VDLAKSRRLSLAHFRHNQINLASHFVMDVFEYKYAKRKLIDVIVIDPPSPFARNK 300

15 Query: 300 KQIPSVTKDYKLLIQALDILTPGQTTIASTNAANLVSCFKKQLSKGFGKASHNYSIQ 359  
 KQIPSV++DY+KLI +ALDIL+P GTIIASTNAAN+TVSQPKKQ+ KGFG ++LQ  
 Sbjct: 301 KQIPSVSRDYHKLITBALDILSPKGTIIASTNAANMTVSQFKQIKNGPSRRPSSMTLQ 360

20 Query: 360 QLPEDFTINKDQGSNYLKVFTIKVR 385  
 QLP DFTIN D++SNYLKVFTIKV+  
 Sbjct: 361 QLPEDFTINKADERSNYLKVFTIKVR 386

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 612

A DNA sequence (GBSx0652) was identified in *S. galactiae* <SEQ ID 1899> which encodes the amino acid sequence <SEQ ID 1900>. This protein is predicted to be minimal change nephritis transmembrane glycoprotein. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -6.85 | Transmembrane | 129 - 145 ( 126 - 152) |
| INTEGRAL | Likelihood = -4.88 | Transmembrane | 48 - 64 ( 46 - 69)     |
| INTEGRAL | Likelihood = -4.93 | Transmembrane | 75 - 91 ( 74 - 97)     |
| INTEGRAL | Likelihood = -4.62 | Transmembrane | 16 - 32 ( 15 - 34)     |
| INTEGRAL | Likelihood = -2.28 | Transmembrane | 163 - 179 ( 163 - 182) |

---- Final Results ----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.3739(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12545 GB:299107 alternate gene name: yetP-similar to  
 hypothetical proteins [Bacillus subtilis]  
 Identities = 299/676 (44%), Positives = 415/676 (61%), Gaps = 33/676 (4%)

Query: 2 KKIKDFASRAINTRLGIFILLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGLL 61  
 KK++ + + +L F +L V+++W KT +Y T+FNLG++ + Q L I NP +  
 Sbjct: 9 KKEVAMKKLFSYKLSFFVLAVILFWAKTYLSTKRNGLGVKGTTORELLIINPFSAYF 68

Query: 62 IIGLALVYKRTAFYITAFITYAIVNILLIANAIYREFSDFITVSAVLASSKTSAGLGD 121  
 +GLAL K K+ I I + + +L AN ++YR F DF+T + S +GD  
 Sbjct: 69 FLGLALLANGRKSAIIMLILDF-LMTFVLYANILYTRFDDFLTFNPKQSGNVG-NMGD 126

Query: 122 SALNLLRIWLDVYVDFILITFLFATKKIHLDDRPPNKRASPSITALSGL-LFSINLFLA 180  
 +++ D+ Y D IILI + + L + KR + S+ LSG+ LF INL A  
 Sbjct: 127 GIFSINGAGHDFIFLDITILLIAVLWRP-ELKEYKMKKRFA-SLVLTSGIALFFINLHYA 184

Query: 181 EIDRPSLLSRGFSNTYIVKALGLSPFSIYSCNTYQAKQRKNGATAQEIAATKAYVABHY 240

-698-

E DRP+LL+R F YIVK LGD +++YY G QT Q +R A++ +L + + Y HY  
 Sbjct: 185 EKDRPQLLRITRFDNRYIVKLGLYNYTYYDGVQTAQTQRAYASSDDLTSEVNYTTSY 244  
 5 Query: 241 AKPNFHYGYIGKRNVMIHLESFQFLIDYKLNIDGKRVVTPFINSLYSKRTVS-PS 299  
 AKWN EY+G KG+N+I IHLESFQ FLIDYKLN G+E VTFP+N L H E V+ F  
 Sbjct: 245 AKPNAEYFGSAKGWNIKHLESFQFLIDYKLN--GEE--VTFPLNKLHGGEVDVTFD 300  
 10 Query: 300 NFFHQVQAQKTSIDARTIMENSLFGLSSGSPMUYGGENTQPAAPHILAQNGYSRAVFG 359  
 NFFHQ GKTSIDAR M+NS+FG L GS V GENT + P IL Q GY+RAV HG  
 Sbjct: 301 NFFHQTCQKTSIDAZLWDSIFGLPGSAFVT- KSENTYQSLPAILDQKQSYTSAVLHG 359  
 15 Query: 360 NVGTFWNRNAYKQGYDYDFDSSYPSKQTKDMSQYGLNDKYMWDASIKYLEHQQFFY 419  
 + +FMR+ YK GYD FPD+S + +N GL DK F +SI LE +QFFY  
 Sbjct: 360 DYKSPWRNDQIYKHIGYDKFPDASTYD-MSDERNVNGKLDKPPFTTESIKLESQAQFFY 418  
 20 Query: 420 TKFITVNGHYPPYTSKLGSEDEGFPPLAKTNDETINGYFATANYLDLALKSFFFLAAQV 479  
 IT++NHYF+ + + A T D T++ YF TA YLD AL+ VF+ LK AG+  
 Sbjct: 419 AHLITLTHNHYFNL---DESDASLKATGDNVDSYPTQATYLDLALBQFFKELKAGL 475  
 25 Query: 480 YDMSIIVMYGDHYGLSNTNPNGLAEELGKDPETWSEYDNAMLRQVPMIHIPOVSKGFTS 539  
 YDMS+I++YGDH GIS N ++ E+LKG+ --ITDYQNAQNRVPMIRVFG-KKGSVN 531  
 Sbjct: 476 YDMSVIMYGDHNGISSENHNRAMKEILGKE---ITDYQNAQNRVPMIRVFG-KKGSVN 531  
 30 Query: 540 NTYGGVDNLEPTLLHILGIDTSKYTGQDLKQDLKSNQVAMRTTQYITPKYNTSYGHL 599  
 +TYGGE+D +PTLLH+ GID+ KY G DL SKD+ VA R G ++TPKY+ +  
 Sbjct: 532 HTYGGVIDVMTLLHLEIGIDSQKYINFQDLFSKDHDDTVAPR-NGDFVTPRYTSVDNII 590  
 35 Query: 600 YTTDSGQETNPDETTKAEIKAIRDAINKOLSTSDSIORGDLRFDENKGLKTVVEKFN 659  
 Y T +G+++ +ET K ++ N+QLS SDS+ DLLRF + N K V+ ++  
 Sbjct: 591 YDTKTGKELKANEET-----KHLKTRVNNQLSLSDVLYKDLLRPFKINDFAVDPSDHY 645  
 Query: 660 YTHSLKALKAKERKLEK 675  
 Y KE++K  
 Sbjct: 646 Y-----GKEKEIK 653

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1901> which encodes the amino acid sequence <SEQ ID 1902>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -6.85 | Transmembrane | 90 - 106 ( 88 - 112)   |
| INTEGRAL | Likelihood = -5.68 | Transmembrane | 146 - 162 ( 139 - 165) |
| INTEGRAL | Likelihood = -4.99 | Transmembrane | 63 - 79 ( 60 - 84)     |
| INTEGRAL | Likelihood = -3.98 | Transmembrane | 178 - 194 ( 176 - 197) |
| INTEGRAL | Likelihood = -0.59 | Transmembrane | 31 - 47 ( 31 - 47)     |

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.3739 (Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

An alignment of the GAS and GBS proteins is shown below:

Identities = 533/713 (74%), Positives = 603/713 (83%)

Query: 1 MKKIDFASRAINTRLGFILLVVIYWLKTIWAYHIDPNGLNSYOLFETIINPIPL 60  
 +KK K + INTRLGI+ L+ YW+KT+WAYHIDF+L L N YQ+ETIINPIPL  
 Sbjct: 16 VKKFKTLITGFINTRLGFITLFCYWIKTWAYHIDPSLDGNIYQVETIINPIPLAF 75  
 Query: 61 LIIGLALVVKRTAFYITAFITAYAVNILLANAIYYRFSDFITVSALVASKTSAGL 120  
 L++G+ALVYK T+APYI +++ Y I+NILL+N+IYYRFSDFITVSA+LASKS AGL  
 Sbjct: 76 LLGVALVYKNTAFYICSWVVYIILNILLNSIYYRFSDFITVSAMLASKSVSAGL 135  
 Query: 121 DSAINLRINLWIVVYDFIILFIFATKCIHLDRPPNKRASFSTIALSGLLFSINFLA 180  
 DSAINLRINW++Y+ DFIIIL L KCI D RPPNKR+A+F+ITALS LL SINFLA  
 Sbjct: 136 DSAINLRINWIIYILDPIILTSIAKKIKNDQRFPNKRAAFATLALSLLLSINFLA 195



-699-

Query: 181 EIDRFELLSSRGFSNTYIVKALGLPSPSIYSGNQTYQAQKRRNGATAQELATKAYVABHY 240  
EIDRFELL+RGFSNTYIV+ALGL+P+I+YSGNQTYQAQKRRNGATA+EL K YV HY  
Sbjct: 196 EIDRFELLTRGFSNTYIVRALGLPAPFTLYSGNQTYQAQKRRNGATAEELIDVTKTVYGHY 255

5 Query: 241 AKPHEFYTGIGGKRNVIHHSVQQQLIDYKINIDGKHEVVTPIINSLYHKSSTVFSN 300  
A P+P+Y+GIGK+H+I++HLESQQQLIDYKINIDGKHEVVTPIINSLYHKSSTVFSN  
Sbjct: 256 AAFDPQTFGIGGKRNVIHLESVQQQLIDYKINIDGKHEVVTPIINSLYHSMATLAFN 315

Query: 301 FFIHQVAGKTSDAKTMENSLPGLSSGSFMVYGGENTQFAAPHILAQNGYSSAVFHGN 360  
FFHQVAGKTSDAET+MENSLPGL+SGSFMVYGGENTQFA P ILAQ GGY+SAVFHGN  
Sbjct: 316 FFIHQVAGKTSDAETMENSLPGLSSGSFMVYGGENTQFATPSILAQGGYTSVAVFHGN 375

Query: 361 VOTFWRNRNAVYQMGWYTFDFSSVTSKQTKNSPQVGLNDKIMFADSIKYLEHMQQPFYT 420  
VOTFWRNRNAVYQMGWY+YTFDFSSVTSKQ NSFPQVGLNDKIMF DSIKYLE HMQQPFYT  
Sbjct: 376 VOTFWRNRNAVYQMGWNTYTFDFSSVTSKQNSKNSFPQVGLNDKIMFDSIKYLEHMQQPFYT 435

Query: 421 KFITVSNHYPTTSLKGESDEBGGPFAKTNDSTINGVFATANYLDTALKSPFEYKAAQVY 480  
KFITVSNHYPTTSLKGES EBGPFPAKT+DSTINGVFATANYLD ALKSPF+YLA G+Y  
Sbjct: 436 KFITVSNHYPTTSLKGESSEBGGPFAKTNDSTINGVFATANYLDAALKSPFYKATGLY 495

20 Query: 481 DNSITVNYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRPVPMIHIPGYSGKFSIN 540  
DNSI V+YGDHYGISN+RN SLA LLGKD ETWSEYDNAMLQRPVPMIHIPGY G I  
Sbjct: 496 DNSITPVLYGDHYGISNRUSSLAFLGKDSSETWSEYDNAMLQRPVPMIHIPGYTNGSIKE 555

25 Query: 541 TYGGSEVNLPTLLHILGIDTSKYTLQGLDLSQNKQVAMRTTGGQYITPKYTNVGHLY 600  
T+GGSE+D LPTLLHILGIDTS+ QGLDLS N Q+VA RT+G Y+TP+VTNYSGLY  
Sbjct: 556 TFGSEIDALPTLLHILGIDTSQVQLQGLDLSQMSQIVAGRTSGTYMPEXTNYSGLY 615

30 Query: 601 YTDGQSEITNDETTKAEIKAIRDATNKLSTSDSISITQDGLLRFDERNGLKTVVEKFNY 660  
T +G EITNDE T A+ K I R A ++QL+ SD+IQTGDLRLPD NGKL ++ +F Y  
Sbjct: 616 NTQTGLEITNDETTIAKTIRESAVQAQLASDAIQTGDLRLPDQGLKAIIDPQFY 675

Query: 661 THSLTALKAKERKLNDRSTSIYSHNNKSTVDLPFAPSYLELQDPNKTHTSK 713  
T LK LK KL STS+YSK ++KST LF APSYLEL TSK+  
Sbjct: 676 TKQLKQLKDIAKLGSESTSLYSHNNKSTQKLFAPSYLELNPVBAADATSE 728

35

A related GBS gene <SEQ ID 8619> and protein <SEQ ID 8620> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop Possible site: -1 Crend: 9  
MoG: Discrim Score: 12.63  
GMA: Signal Score (-7.5): -2.99  
Possible site: 30  
>>> Seems to have an uncleavable N-term signal seq  
ALOM program count: 5 value: -6.85 threshold: 0.0

45 INTEGRAL Likelihood = -6.85 Transmembrane 129 - 145 ( 126 - 152)  
INTEGRAL Likelihood = -4.88 Transmembrane 48 - 64 ( 46 - 69)  
INTEGRAL Likelihood = -4.83 Transmembrane 75 - 91 ( 74 - 97)  
INTEGRAL Likelihood = -4.62 Transmembrane 16 - 32 ( 15 - 34)  
INTEGRAL Likelihood = -2.28 Transmembrane 163 - 179 ( 163 - 182)

50 PERIPHERAL Likelihood = 3.76 103  
modified ALOM score: 1.87

\*\*\* Reasoning Step: 3

55 ----- Final Results -----  
bacterial membrane --- Certainty=0.3739 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

45.2/63.1% over 643aa

Bacillus subtilis

65 BGAD[107893] hypothetical protein Insert characterized  
GP[2116767|dbj|BAA20118.1||D86418 YfnI Insert characterized

-700-

GF|2633039|emb|CAB12545.1||Z99107 alternate gene name: yetP-similar to hypothetical proteins Insert characterized

PIR|D69815|D69815 conserved hypothetical protein yfnI - Insert characterized

5 QRF00125(286 - 2280 of 2742)  
 RGAD[1078]||[80726 (3 - 646 of 653) hypothetical protein { acillus subtilis} GP[2116767][dbj]  
 AA2018.11||D6846.8 YfnI { acillus subtilis} GP[2633039][emb|CA 12545.1||Z99107 alternate gene  
 name: yetP-subtilis] conserved proteins { acillus subtilis} PIR[D69815][D69815 conserved  
 10 hypothetical protein yfnI - acillus subtilis  
 %Match = 28.5  
 %Identity = 45.1 %Similarity = 63.1  
 Matches = 297 Mismatches = 227 Conservative Sub.s = 118

36 66 96 126 156 186 216 246  
15 EVVKKRPSLRIDLTVKKVEPTG\*LNWYONLFFPVTEHLI\*FFRFRONSIVYS\*TVL\*QIFIFHTTEEDLSLSEYTKVEV

20

|  |                                                                           |     |     |     |     |     |     |     |
|--|---------------------------------------------------------------------------|-----|-----|-----|-----|-----|-----|-----|
|  | 276                                                                       | 306 | 336 | 366 | 396 | 426 | 456 | 486 |
|  | II*SEILSGKGLKVSVTFYKKIDFASRINRLGIFILLVVIWKLTIWAHTDPLNGLSRYOLFITIINIPIPLLI |     |     |     |     |     |     |     |
|  | :                                                                         | :   | :   | :   | :   | :   | :   | :   |
|  | :                                                                         | :   | :   | :   | :   | :   | :   | :   |
|  | MNEELKVFVKVEVAMKLVLSYKLSFFVLAVLFWKTYLLSYKTEPNFLGVKGTTQELLIPFPSSAVFF       |     |     |     |     |     |     |     |
|  | 10                                                                        | 20  | 30  | 40  | 50  | 60  |     |     |

25

516 546 576 606 636 666 696 726  
IGLALVVRKTYAFITATITVAIVNILLIANAIYYREFSDFITVEAVLASSKTSAGLGDSALIRMLRWLDVYVDFDIIL  
: ||| : | : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||  
LGLALLAKGKRSALIMLIDF-LMTFVLVYANILFYRFDFDLPFNIL-KQSGNVNMGSGIGFISNAGHDIYTFPLDILIL  
80 90 100 110 120 130 140

[illegible][illegible][illegible][illegible]

55  
 1710    1740    1770    1800    1830    1860    1890    1920  
 NYLDALAKSFFFYKAGGVDSITVMYGDHYGISNTRNPSLAELKKDPTWSSYNNMLORVPMIHIPIGYSKGFISN  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 RYLDEALSQFFFKKAGLYDSIMTMDHNGHSINENRMKILTK---ETTDYNACNQVRPLMIRVPG-KKGVDNH  
 470                 480                 490                 500                 510                 520                 530

[illegible]

65            2190            2220            2250            2280            2310            2340            2370            2400  
AIRDATNQGLSTSDSIQTGDLIRFDENNGIKTVVEKEFNTHSLKALKAKERKLRKSTSTYSKHNNKSTVDLFHAPSYL  
::    |:|:| |:|:| :|:|:| :|:| :|:|  
NLKTRVNGQLSLSDSVLYKDLIRPHKLNDFKAVDPDSYHYGKEKEIK

-701-

620 630 640 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 613

A DNA sequence (GBSx0653) was identified in *S. agalactiae* <SEQ ID 1903> which encodes the amino acid sequence <SEQ ID 1904>. This protein is predicted to be 50S ribosomal protein L20 (rpIT). Analysis of this protein sequence reveals the following:

```

Possible site: 37
10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3392(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9387> which encodes amino acid sequence <SEQ ID 9388> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GF:CB14845 GB:Z99118 ribosomal protein L20 [Bacillus subtilis]
 Identities = 70/89 (78%), Positives = 78/89 (86%)

Query: 1 MFRTAKEQVMNSYYAYRDRQKKRDFRKLWITRINAAARMNGLSYSQIMHGLKLAETSV 60
 +++ A +QVM S YA+RDRQKKRDFRKLWITRINAAARMNGLSYS+LMHGLKL+ IEV
25 Sbjct: 31 LYKVANQVMKSNYAFDRQKKRDFRKLWITRINAAARMNGLSYSRIAMHGLKLSGIEV 90

Query: 61 NRKMLADLAVNDAAAFALADAAKALGK 89
 NRKMLADLAVND AF LADAAAL L K
30 Sbjct: 91 NRKMLADLAVNDLTAFLADAAKALGK 119

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1905> which encodes the amino acid sequence <SEQ ID 1906>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.06 Transmembrane 94 - 110 (94 - 110)

----- Final Results -----
 bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 87/89 (97%), Positives = 88/89 (98%)

45 Query: 1 MFRTAKEQVMNSYYAYRDRQKKRDFRKLWITRINAAARMNGLSYSQIMHGLKLAETSV 60
 +FRTAKEQVMNSYYAYRDRQKKRDFRKLWITRINAAARMNGLSYSQIMHGLKLAETSV
Sbjct: 31 LFRTAKEQVMNSYYAYRDRQKKRDFRKLWITRINAAARMNGLSYSQIMHGLKLAETSV 90

Query: 61 NRKMLADLAVNDAAAFALADAAKALGK 89
 NRKMLADLAV DAAAFALADAAKALGK
50 Sbjct: 91 NRKMLADLAVDAAAFALADAAKALGK 119

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 614**

A DNA sequence (GBSx0654) was identified in *S.agalactiae* <SEQ ID 1907> which encodes the amino acid sequence <SEQ ID 1908>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.64 Transmembrane 32 - 48 (32 - 48)
 INTEGRAL Likelihood = -0.32 Transmembrane 3 - 19 (3 - 19)

----- Final Results -----
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 615**

20 A DNA sequence (GBSx0655) was identified in *S.agalactiae* <SEQ ID 1909> which encodes the amino acid sequence <SEQ ID 1910>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-12.63 Transmembrane 747 - 763 (743 - 772)
 INTEGRAL Likelihood =-12.52 Transmembrane 840 - 856 (835 - 856)
 INTEGRAL Likelihood =-11.20 Transmembrane 447 - 463 (440 - 466)
 INTEGRAL Likelihood = -5.79 Transmembrane 351 - 367 (346 - 372)
 INTEGRAL Likelihood = -4.25 Transmembrane 517 - 533 (516 - 537)
 INTEGRAL Likelihood = -1.49 Transmembrane 397 - 413 (396 - 413)
 INTEGRAL Likelihood = -0.96 Transmembrane 799 - 815 (799 - 817)

----- Final Results -----
bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

35 A related GBS nucleic acid sequence <SEQ ID 9349> which encodes amino acid sequence <SEQ ID 9350> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
[Archaeoglobus fulgidus]
Identities = 100/483 (20%), Positives = 210/483 (42%), Gaps = 61/483 (12%)

Query: 351 LFFPIILYIAAALVILTLTWTFVEBERTNAGILKALGYSURQVIFKPIIYGFAGLTGTL 410
LPP LV+ +T ++R + N +++ALG++ +++ ++Y +G +T

Sbjct: 276 LFFAFFIILVSIWYLLSRIPLQLGNAVMRALGFTIRNEIMLHYLQXPLMGFFASTA 335

Query: 411 GIIGGHYLLPRIISDIISKDLTINTQYHLFINYSLLAFVFSLSISIVLPVFVI----- 463
G++G++ +S I+ L+P L L+ +L+ +P++

Sbjct: 336 GIIVAGFPASQQLTSQYIT-PLNLGPTYSKPHLEVYSLSMAGLITLPTISGFINAYQASRV 394

Query: 464 ----TRRELKEKAFFLLPKPPARGSKTALRYINWIWKKLSFTQKVTARNIPRYKQRM 519
R EAA ++ A S+I W ++ ++ RNIFR K+R ++

Sbjct: 395 DIVKALRGYAEVAAVSSTARIDALFSRI-----W---RMRLKPLKALRNIFRSKRRTAT 445

Query: 520 TTFGVAGSVALLSGLGQSILKQTVNENFCRIMPYDILLTYNTNAPPKLLELLSKDSK 579

```

-703-

```

 +IF + +L+ + + S + FG++ YDI ++ E+L K K
Sbjct: 446 SIPSI VACTSLILASMVFDSPDYVWQLQFGKVYAYDIKVSLEGYDGK----EVLEKVRK 501

5 Query: 580 IDKY-----QPHLENLNDESIPGQINKQSISLPTDKKQLPFIYLAQTATNKSIHL 631
 +D PI++E E++P +L I Q L +Y E +
Sbjct: 502 MDGVLFAPFAVEMPIYVEKGEAVP-----TLLASNFQTLINNVNABG----EKLI 549

10 Query: 632 INKGIILSKLQAFYHIVAGDPFHL-----SHSQTLPSRKLKITGVNANVGHYIWTK 685
 ++GII SK + + G+ + + + + + V A++
Sbjct: 550 PSEGIIPSKTAMGNLSLVEGEKVSYPTEPGKLEAEVDVEMIFLLSVATASL----- 601

15 Query: 686 QYRTIFPKKAKDAPLVKLTAKHIANNAEKLLEINGVESLQNALQLASVPAVRSID 745
 Y+ I + N +V + +IA +AEK+ ++GV+ ++ S+E ++
Sbjct: 602 DYPSTRISGVDG-PNRIVVDADGRIA-EIAEKIRQMDGVKKVSTVIRAQSIHELNGFPY 659

20 Query: 746 GSKTILVVVSLLLAIVILYNLTINILAEKRELSITKVLGFYNEEVLTYIYRETIILSTI 805
 + + + L +N T+I++ ER REL+T++LG+ + E+ + + E + ++ +
Sbjct: 660 AFIAPSLFPVSLGFPAVPTTTSISVIERSEKELATLMLGYTSEIILSILENLFPAIL 719

20 Query: 806 GVI 808
 G++
Sbjct: 720 GLV 722

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1911> which encodes the amino acid sequence <SEQ ID 1912>. Analysis of this protein sequence reveals the following:

```

 Possible site: 34
>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -14.33 Transmembrane 749 - 765 (739 - 775)
 INTEGRAL Likelihood = -10.88 Transmembrane 845 - 861 (834 - 865)
 INTEGRAL Likelihood = -6.64 Transmembrane 350 - 366 (344 - 369)
 INTEGRAL Likelihood = -6.53 Transmembrane 22 - 38 (19 - 42)
 INTEGRAL Likelihood = -6.32 Transmembrane 520 - 536 (515 - 537)
 INTEGRAL Likelihood = -4.99 Transmembrane 446 - 462 (445 - 465)
 INTEGRAL Likelihood = -2.92 Transmembrane 396 - 412 (395 - 413)
35 INTEGRAL Likelihood = -0.80 Transmembrane 800 - 816 (800 - 819)

----- Final Results -----
 bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AA869436 GB:AB000977 A. fulgidus predicted coding region AF1820
 [Archaeoglobus fulgidus]
45 Identites = 101/542 (18%), Positives = 237/542 (43%), Gaps = 42/542 (7%)

 Query: 350 IFPVVLYLVAALVAFPTMTFRYVDEERTSSGLLKAIGVSNKDI SLKFLIYGLLASFIQTTL 409
 +FP LV+ + + +R + + +++A+G++ +I L +L Y LL F +T
Sbjct: 276 LFPAPFPLVSLPMFYALLSRIFRIQLGNTIAVGRALGPTREINLHLYQYPLMGFFASTA 335

50 Query: 410 GIIGQTYLLSTLISRLITGA---LTIGKTHLYSYWFGNGIAYLLAMSAVLPAYLIVKKE 466
 G++ G + L S+ +T + K HL Y L +S L AY + +
Sbjct: 336 GLVAGFASQLTSQYITFLNLPIYVYKPHLEVSLSIMAGTLTPTISGFINVYQASRVD 395

55 Query: 467 LFLN-----AAQLLLPKPPSKGAKIWLHSLTFVWKALESPTHKVTIERNIPRYKQRMLMT 519
 + AA + + + ++IW L F ++ +RNIFR K+R ++
Sbjct: 396 IVKALRGYAEVAVSPFIARIDALEPSRIWRMLIF-----RIALENIPRSKRKSLT 446

 Query: 520 IVCVAGSVALLFAGLGIQSLAKVVEHQDGLTITYDILAVGSNAKATATEQTDLASTYLKQE 579
 I + +L+ + S V++ QFG + YDI + L Y +E
60 Sbjct: 447 IPSIVACTSLILNBMVFDSPDYVWQLQFGKVYAYDI-----KVSLEGYDGKE 494

 Query: 580 PITGYQKVSYASLITLPVKGILP---DKQSISILSSS-ATLSLPYFNLLDSQKQKVPPTS 635
 + +K+ P +P +K ++ + A++ +N+ +++ +K IP+
65 Sbjct: 495 VLEKVRKMDGVLFAPFAVEMPIYVEKGEAVPTLLASNFQTLINNVNABGSL--IPSE 552

```

Query: 636 GVLISEKLASYKVKGDQGLWIDRKGQSYKVTIKQVIDMTVGHYILMSDITVFKNHFQGL 695  
 G++ S+ + G+++ + G+ ++ ++ L+ T ++F +  
 5 Sbjct: 553 GIIPSKTAHKLNSLVSGKGVSYTFGK-----LEAEVEDVEMIPFLSVATASLDFSR 607

Query: 696 EAPAYLIKVDKDSKNIKGTASDLTLKAIKRAVSGQVNMIIKSVQLVVTSLAQVMTLLVF 755  
 + V D D I R A + + ++ V5 + +S++ + ++F  
 Sbjct: 608 SGVDGFNRIVVDADBEIARIAEKIRQMDGVKKVSTVIEAQESIEELMGFFYAPTAFSLF 667

Query: 756 LSILLAVILNLTINIAARRIEHLSITKVLGYDQEVLYTYRETISLGVLLIGLYL 815  
 + L ++N T+I++ RR RRL+T+++LG+ +E+ + + E + +++G++  
 10 Sbjct: 668 FGVLGTFAAVNTTSISVIRRGRLATLRLGLTTSREIIIGLILENLFVAIGLVLFALEI 727

Query: 816 GKGLSTYIMINISTGDIQFGVKNDAYVYLVPIVLVLSLLAVLSIWNHRLKLVMDLEALK 875  
 + + + + + + + + +L + ++ + + R + ++D+ + K  
 15 Sbjct: 728 AYSTAYFFPFGSEELYIMFWIYPTFAATVLAVFATILLALLPARRVSEMDIAVKTK 787

Query: 876 SI 877  
 I  
 20 Sbjct: 788 EI 789

An alignment of the GAS and GBS proteins is shown below:

Identities = 377/857 (43%), Positives = 543/857 (62%), Gaps = 7/857 (0%)

Query: 3 KTFWKDIYSITTSKGRFSSILLMMGLSPAFILGKVSAPNMQRQAQYLAHHHVMIDTV 62  
 K+ WKDI R+I SKGRF S+ LM LGSFA +GLKV+ P+M+RTA YL H VMD+TV  
 Sbjct: 4 KTWKQILRAIKNSKGRFISLTFMGLGSPALVGLKVTGPMERTASRYLERHQVMDLTV 63

Query: 63 FNSWGLKIDQTVLESLSGQVEPSYFVDTTQQNSKSYRLYENTKTISTPDLVKRLFL 122  
 S + D+ L+LKG+ +E+ +D+ N KS RLTS K+S LVKG P  
 30 Sbjct: 64 LASHQFSQADQQLDLTGALHLEYGHLLDVSLTNSKQSLRLYSVPKKSVKPVLYVSGSEWK 123

Query: 123 NKSEIALSPQRNKYIAIGDKINFPQDNKLPSVTGPIITGVFNSGTWIKNLQSGGTG 182  
 ++++ LS K Y IGD++ L + T +VGF NG+EBNK+NLGSG TG  
 35 Sbjct: 124 RETDLVLSSSLANQYIGDELAVTSPMGGLTITTH-PQVAFANGSGSEWKNLQSGSGTG 182

Query: 183 DCDLSYGVLDKTAHPSPVYTMARVTFKDLRLINPFSISYKEKVAKYQKVKRKLHINK 242  
 DG L+Y ++ F S ++R+ F LRL N FS Y++V + Q + L ++  
 40 Sbjct: 183 DGSLYAYAVNPNVPS-AFNLRLIRFSLRLTNAPSKDYQKRVTONQAHLNLLKQNG 241

Query: 243 IRYTKKESLRKIDEEKSLIAQKQINRLNDSLAMPLOQAQIMKIKQDLGLLKR 302  
 RY ++ + + +L K ++ + + + S Q + +I+Q + +L K  
 45 Sbjct: 242 KRYDDLQNYDLALINGRAALAKETVLAASENLTFLEGSALQEAHQIEQKQALAE 301

Query: 303 TKELLKLRHNTQIMSEPOIIVNNTTFPGQGYNTFDSSNTNSTKISINLFPILLYLVAAL 362  
 K+L +++ +E P + YNR+T DGS+GY+T+ +ST S S + N+FP+LYLYLVAAL  
 50 Sbjct: 302 EKQLBQVQNTDKLEKPSYLTNRSTLPGSGBYHYATYTTSTISNWNIPFPVLYLVAAL 361

Query: 363 VLTMTWTRFVEEHTRAGILKALGYSDRQVFPKFIIVGFIAGLTGTLIGGHYLLPFI 422  
 V TMTWRA+V+ERT+G+LKA+GYS++ + KP+IVG +A LGTTIGIIG YLL +  
 55 Sbjct: 362 VAPTTWTRVDRERTSGLLKALGYENKDISLFLIYGLLASPLGTLTIGSTYLLSTL 421

Query: 423 ISDLSKDLTINTQYHLEFNLSLAFVFLSLIVLPVFTVITRELEKKAFLLLPKPFA 482  
 IS+I++ LFI T + + Y+ +A++ +LS VLP +I+ +EL AA LLLPKP+  
 60 Sbjct: 422 ISEILVGLTIGKTHLYSYWFGNGLAYLLAMLSAVLPAYLIVKLEFLNAGLLPKPFS 481

Query: 483 GKSKIALEYINWIKKLSFTQKVTARNIPRYKORMIMTIPGVAGSVALLPSGLIGQSSLE 542  
 KG+KI LE++ ++WK LSPF KVT ENIPRYKQRM+MTI GVAGSVALLP+GLIGQSSLE  
 65 Sbjct: 482 KGAKIWEHLTFVWKALSFTHKVTIENIPRYKQRLMTIVGVAGSVALLPAGLIGQSSLA 541

Query: 543 QTVNEHFGRLMPYDILLTYTNASPPKILKLS--KSKIDKYQPHLEKSDSEIPGGIN 600  
 + V PG + YDIL + A+ + +L S K I YQ + +L + G +  
 70 Sbjct: 542 KVVEHQPGDLTTYDILAVGSAKATATEQTDLASLKGPEITGYQVSYASLTLVKGSD 601

Query: 601 KQISLFTIDKQLLPFIYLRATINKSLHNNKGLIIKKLAQFHYHNGDFHLSHQ 660  
 KQIS+ + L P+ L ++ K + + G+IS+KLA +Y V GD + L+ +  
 75 Sbjct: 602 KQISILSSSATSLSFYENLJASQEQKVKVPIPSGVLISEKLASYKVKRGDQVLAITURK 661

-705-

Query: 661 TLPGRKLKITGVNANVCHYIPMTKQYYRTIFKKEAKDNAPLVCL--TKHKIANNLAEKL 718  
 S K+ I V++ VGHY+ M+ Y++ FK A+L+K+ K A L  
 Sbjct: 662 G-QSYVTIKQVIDMTFVGEYILMSDTYFKRHFKGLEADPAYLIKVKDKDSKII KETASDL 720

Query: 719 LKINGVBSLTQNALQLASVBAVVRSLDGSMTILVVSLLLAIVILNNINMLAERKREL 778  
 L + + +++QH + SV+ VV SL+ NT+LV +S+LLAIVILNLT IN+ASR REL  
 Sbjct: 721 LTKAIRAVSQNVHILKESVQLVVTSINQVMTLLVFLSLILAIVILNLTINIAERREL 780

Query: 779 STIKVLGFPYNEVTLIYRETILSLTIGVILQTSITLHQRQMLLIGSDQILPQEKVSP 838  
 STIKVLGFY++SVILYIYRETI LS +G++LG G LH +M +I + I PG KV  
 Sbjct: 781 STIKVLGFPDQEVTLIYRETISSLVSLGLGLIYLGKSLHYIMTISTGDIQFQVEVDA 840

Query: 839 TFIIPISVVVILQKL 855  
 +++PT V++ +L L  
 Sbjct: 841 YVLLVPILVILSLAVL 857

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 20 Example 616

A DNA sequence (GBSx0656) was identified in *S.galactiae* <SEQ ID 1913> which encodes the amino acid sequence <SEQ ID 1914>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2757 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB89431 GB:AB000977 ABC transporter, ATP-binding protein  
 [Archaeoglobus fulgidus]  
 Identities = 112/230 (48%), Positives = 167/230 (71%)

Query: 4 IEMKHSYKRYQTGETEIVANDISFSIERGELVVLGASGAGKSTVLNVLGMDNSSEGE 63  
 + ++ +K YQ G E+ A I+ IERGE +V+LG SG GK+T+INT+GG+D + G  
 Sbjct: 2 LRLIEDVWVYQVQVKSALRGINLEIERGEFNVVLGPGSGCKTMTLNIIGIDRPTGR 61

Query: 64 VLIDGKNIANITYRELTTRYRYDVGFVFOFYNVNLTALENVELASEIVPKALDAQOAL 123  
 V+ DGG+I NY LT +RR +VGF+PQF+NL+P LTA ENVE+A+++V D + L  
 Sbjct: 62 VIFGDKDITNYNEDRLTYHRNRNVGFIPOFFENLPTLTARENVEIADLVSPRDVDEVL 121

Query: 124 ENVGLGHRINHFPAQLSGSGEQQRVALARAIKPKLLLCDEPTGALDYQYQKQVAILQK 183  
 + VGL R HFPA+LSGSGEQQRVALARA+ K P +L DEPTG+LD++TGK VL ++++  
 Sbjct: 122 KMGGLADRAEHFPAELSGSGEQQRVALARLVKNPPTILADBPVGLSDFETGKAVLKVME 181

Query: 184 MAQSKETTVIIVTNTALAPTANRVIHMDSKISDIVINENPSDIQNTLEY 233  
 + + + T ++VTIN+A+A EA+RV++ D K+ + N +P+D I++  
 Sbjct: 182 INRKEGITFVIVTNTSATAADRVVYLRDGVKVERVETNHPADPDRIQW 231

There is also homology to SEQ ID 1354.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 617

A DNA sequence (GBSx0657) was identified in *S. agalactiae* <SEQ ID 1915> which encodes the amino acid sequence <SEQ ID 1916>. This protein is predicted to be DNA topoisomerase I (topA). Analysis of this protein sequence reveals the following:

```

5 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.4716 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9821> which encodes amino acid sequence <SEQ ID 9822> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CB13465 GB:299112 DNA topoisomerase I [Bacillus subtilis]
Identities = 442/690 (64%), Positives = 535/690 (77%), Gaps = 10/690 (1%)

20 Query: 27 LVIVESPAAKATIEKYLGRNYKVAVSGHIRDLLKSSMSIDFENNYEPQVINIRGKGPLI 86
 LVIVESPAAKATIE+YLG+ YKV AS+GH+RDL KS-M +D E N+EP+YI IRGKGP++
 Sbjct: 5 LVIVESPAAKATIERVLGKKYKVKASMGHVRDLKPKSMGVVDIEQNFPKYITIRGKGPLV 64

Query: 87 NDLLKEAKKAKKVYLAADPDREGEAISWHLAHLIDLDKEDRMVRVFNITKDAVGNAPVE 146
 +LK AKKAKKVYLA+DPREGEAI+WHLAH LDLD RVVFNITKDA+K +F
25 Sbjct: 65 KELTKAANKKAKKVYLAADPDREGEAIWHLAHLIDLDNSDCRVVFNITKDAIKESFKH 124

Query: 147 PRQINMDLVDAQARRVLDRIWGVYSISPILMKVKVGLSAGKQVQSVALKLIIDRENSIKA 206
 PR INMDLVDAQARR+LDR+VGY ISPILMKVKVGLSAGKQVQSVAL+LIIDRE BI
 Sbjct: 125 PRQINMDLVDAQARRILDRLVGYKISPILMKVKVGLSAGKQVQSVALRIIDREKIND 184

30 Query: 207 PQPEYWTIDGSPFKGTRKFNATFYGLDGKKFKLNNEDVKTVLNRIKTIIDFLVEKVEKK 266
 F+PEYWTIDG+F KS F A+F+G +GKK L+ DVK +L ++K +++ VEKV KK
 Sbjct: 185 FKPEYWTIDGTLFGKQTFEASFPKNGKKFLNLSADVKEILSLEKGNQITVEKVTKK 244

35 Query: 267 EERRNDLPFTTSS+QQDAANKINPFRKTHMIAQLYBGLSLGTAGHQLITMYRDTST 326
 ER+RN LP+TTS+LQQ+AA K+NR +KTHMIAQOLYB+ IG G GLITMYRDTST
 Sbjct: 245 ERKRNALPFTTSTLQQAARKLAFRAKTHMIAQLYBGLSLGTAGHQLITMYRDTST 304

40 Query: 327 RISPLAQNEATEFTFTRPGANYSKHGNK-VIWSGAQQAHEAIRPSSVHTPESIAKYLD 385
 RIS A+EA FI +G + K K AQDAHEAIRP+SV P + L
 Sbjct: 305 RISMTAVDEAAAFIDQTYGKEFLQGRKPAKCNHQAQDAHEAIRPSSVLAKEPSELAVLG 364

Query: 386 KDQLGLATLINNRFIASQNTAVFDTMKNVLTQNGVTPPIANGSQVKPDGMYAVND- 441
 +DQ+LY LIR RF+ASQM AV DTM V+LT NG+TP ANG+VKF G+M VY +
45 Sbjct: 365 RDQGRLYKLIWRFVASQMAPAVLDTMSVDLTNGLTFRANGSKVPSGFMKRVVSGKDD 424

Query: 442 --TDKNMLPMEBEGESVKKVMNTPHQHTOPPARPSESLIKTLENGVGRPSTYAPTL 499
 +K+MLPD+EG+V + PECHPTOPP R++EA L+KLEK G+GRPSTYAPTL
 Sbjct: 425 QMEKORMLPDLQKEDTVLSKDIEPEQHPTQPPRYTRARLVKTLBERGIRPSTYAPTL 484

50 Query: 500 ETIQGRYVYVLAARKEPTELTGEIVNSELIVEFPDVIDVPTTAMAGKLDVEVIGKQKQ 553
 +TIQAR YV L KRF FTGLG+TV LI+EFPF+I+V PTA+ME LD VE G +M
 Sbjct: 485 DTIQRGYVALNKRFPVTELGQIVLIDLIMEFPPEIINVEPTANMERDLDHVBSGNTENV 544

55 Query: 560 KIIDEFYKPFKEKELAKAETEMEKIQLKDEPAGFDCELCSPMWIKLGRYQKFPVACSNFPE 619
 KIID FY PEK + KAR+EM+++++I+ E AG DCELC SPW K+GRYQK ACNFP+
 Sbjct: 545 KIIDNFTTDPEKRVKAGSEMEVEIEPVSAGDCELCSSPMVYIMGRYQKFLACSNFPD 604

60 Query: 620 CHNTKAITKEIGVIPCIOQKGVIERKTKRNRIFYGCDRYPCEFTSNDPIKRTCKPKN 679
 C NTK I K+IGV CP C +G ++ERK+K+ R+FYGCDRYP+CFE SWDKP R CPK
 Sbjct: 605 CHNTKPIVKGIVKCPSCGEGNIVERKSGKKRVFYGCDRYPCEFTSVNDKPIERKCKPKG 664

```



-707-

Query: 680 DFLVEKKVREGGKQVVCSENEKCDYQSEKIK 709  
 IAVEKK++ G QV C +CDY+RE K  
 Sbjct: 665 KMLVEKKLK-KIGIQVC--VECDYKEEPQK 691

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1917> which encodes the amino acid sequence <SEQ ID 1918>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5445 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below:

Identities = 595/704 (84%), Positives = 656/704 (92%), Gaps = 1/704 (0%)

- Query: 6 TTTKTSKTKTSKKKSATAKDNLVIVSPAPAKTIEKYLGRNFKVVASVGHIRDLKKSSMS 65  
 T KT TKK++ KK +TAKDNLVIVSPAPAKTIEKYLGR+YKVASVGHIRDLKKSSMS  
 20 Sbjct: 7 TKPKTGTSTTKTKKSTAGNLVIVSPAPAKTIEKYLGRSYKVASVGHIRDLKKSSMS 66
- Query: 66 IDFDNNVEPCQINIRGKGPLINDLKRAKKAKVYVLASDPDRBGEALISWHLAHLIDLDKE 125  
 IDP+HNYEPCQINIRGKGPLIN LKRAK AKKVYVLASDPDRBGEALISWHL+HIL LD +  
 Sbjct: 67 IDFDNNVEPCQINIRGKGPLINSLKKEAKAKKVYVLASDPDRBGEALISWHLAHLIDLPQ 126
- 25 Query: 126 DNNRVVFNEITKDAVKAFVPEPQI+MDLVD+QARRVLDRIVGYISPLMKVKVKGLS 185  
 D NRVVFNEITKDAVK+AFVPEPQI+MDLVD+QARRVLDRIVGYISPLMKVKVKGLS  
 Sbjct: 127 DNNRVVFNEITKDAVKAFVPEPQIDMDLVD+QARRVLDRIVGYISPLMKVKVKGLS 186
- 30 Query: 186 AGRVQSVALKLIIDRENLIKAFQPEVWITDGSFKKTKFNATFYGLDGGKPKLNNRD 245  
 AGRVQSVALKLIIDREN+IKAF +FVW+IDG FKETK+FK ATFYG+GKK KL NN D  
 Sbjct: 187 AGRVQSVALKLIIDRENLIKAFVPEYWGIDGLFKKGTKKFQATFYGNGKTKLDNNRD 246
- 35 Query: 246 VKTVLKRIKTDPELVKVEKKEKRRRNALPYTTSLQDDAANKINFRTRKTMTAIAQLVE 305  
 VK VL ++ ++FLV KV+KKEKRRRNALPYTTSLQDDAANKINFRTRKTMTAIAQLVE  
 Sbjct: 247 VKEVLAKLTNEDPLVSKVDKKEKRRRNALPYTTSLQDDAANKINFRTRKTMTAIAQLVE 306
- Query: 306 GLSLGTAGHGLITVMRTDSTRISPLAQNEATEFITRFGANYSEKGNKVKVAGQAQAH 365  
 G+ LG G QGLITVMRTDSTRIS+AQN+A +FI NRFGANYSKGN+VHN SG QDAH  
 40 Sbjct: 307 GIHLGENTQGLITVMRTDSTRISFVQNDAAQFIINRFGANYSKGNRVHNTGOVQDAH 366
- Query: 366 EAIRPSSVNHTPSIAKYLKDKQLKLYTLIWNRFIASQNTAAVDFIMKVNLTQNGVFTFA 425  
 EAIRPSSVNHTP+SIKYL+KDKQLKLYTLIWNRF+ASQNTAAVDFIT+KVN QNGV F+K  
 45 Sbjct: 367 EAIRPSSVNHTPDSIAKYLKDKQLKLYTLIWNRFVASQNTAAVDFITVKNLEQNGVIFVA 426
- Query: 426 NGSQVKFDGYNAVYNETDKNKMFLPMDEBESGVKKVNFNPQHTQPPARFSEASLITKLE 485  
 NGSQ+KFDGYNAVYNE+DKNKMFL+M BE+VKK++T+PEQHTQPPAR+SEA+LITKLE  
 Sbjct: 427 NGSQMKFDGYNAVYNE+DKNKMFLPMASBETVKKISTSPBQHTQPPARFSEATLITKLE 486
- 50 Query: 486 ENGVORPSTYAPTELTQKRYTVKLAAKRFSETELGEIVNSLIVEFFDIDVDVPTAESE 545  
 ENGVORPSTYAPTE LQ+RYTVKL+AKRFSETELGEIVN LIVEFFDIDVDV PTAESE  
 Sbjct: 487 ENGVORPSTYAPTELVIORRYTVKLAAKRFSETELGEIVNKLIVEFFDIDVDVPTAESE 546
- 55 Query: 546 GKLDVEIGKBQWQKIDIEFYKPFPEKLAARTEMEKIQIKDEPAGFDCLOSGSPWIKL 605  
 GKLD+VEIG+BCQW+ID+FY+PF KEL KAE+E+EKIQIKDEPAGFDC++CG PWIKL  
 Sbjct: 547 GKLDQVEIGKBQWQINIDQFYOPFVGEINKARSEIEKIQIKDEPAGFDCDVGSPHWIKL 606
- Query: 606 GRYGKFYACSNFPECHNKAITKEIGVIPCQKQGVIERKTKRNRIFYGCDRYPECEFT 665  
 GR+GKFYACSNFPEC NTKAITKEIGV CP+C KQVIERKTK+NRIFYGCD+Y+PECEFT  
 60 Sbjct: 607 GRGKFYACSNFPECHNKAITKEIGVITPCVCHKQGVIERKTKRNRIFYGCDQYDPCSF 666
- Query: 666 SWDKPIGRCTCPKSNDFLVEKKVREGGKQVVCSENEKCDYQSEKIK 709  
 SWD PIOR CPKS D+L+EKKVR GKGQV+CSNE CDY+SEKIK  
 Sbjct: 667 SWDLPIGRACPKSGDYILIEKKVR-GGKQVMSNETCDYQSEKIK 709

-708-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 218

A DNA sequence (GBSx0658) was identified in *S. galactiae* <SEQ ID 1919> which encodes the amino acid sequence <SEQ ID 1920>. Analysis of this protein sequence reveals the following:

Possible site: 43  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2578 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP: AAD35341 GB: AR001708 DNA processing chain A [Thermotoga maritima]  
Identities = 97/231 (41%), Positives = 149/231 (63%), Gaps = 2/231 (0%)  
Query: 51 FIENYKQLDLKKLRQEFKKPPV--LSILDENYPLELKEIYNPPVLLFYQGNIELLSKPKL 108  
P+R + +L+ + + +K V +S + +Y L+EI PP +L+ +G+ ELL + +  
20 Sbjct: 41 PLEKQKKEELERQKELIRKHNKLVSWEDDYPQHLREIRYPPAVLFRGDAELLKECKV 100  
Query: 109 AVVGARQASQIGQCSVKIKIETNNQFVIVSGLARGIDTAHVSAKNGGSSIAVIGSGL 168  
VVG R+ + G K+ +K + FVIVSG+A GID+ AH AL +GG + +AV+G+G+  
30 Sbjct: 101 GVVGTRRPTSYGVNVTKRFVKLLSEYFVIVSGMAFGIDGVAHKEALSSGSKTVAVLGTGV 160  
Query: 169 DVYYPTENKQLQEYMSYNHLVLSEYPTGEPLKPHFPERNRIIAGLCOQIVVAEAKNRSG 228  
DV YP N+ +L + N V+SEY G+ K HPP ENRIIAGL I+V EA + +GG  
40 Sbjct: 161 DVVYPERNRLFHEIVNGCVVSEYPMGTARKHHPFARNRIIAGLSDAITVETAKISG 220  
Query: 229 SLITCERALEBGRVFAIPGNIIIDKSGDCHILQEGAKCIISGKILISEY 279  
+LIT + ALE GR+VFA+PG+I S+G + +LI+ GA + +D+ + +  
50 Sbjct: 221 ALITVKFALESQRDFAVFGDIDRKTSEGTNYLRSYGAYPLTDEEDLSTHF 271

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1921> which encodes the amino acid sequence <SEQ ID 1922>. Analysis of this protein sequence reveals the following:

Possible site: 50  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2855 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

45 Identities = 185/279 (66%), Positives = 238/279 (84%), Gaps = 1/279 (0%)  
Query: 1 MNHFELFKLKGALTNLNLHNIINYLKQNSLTSLSVRNMAVVSCKGNPTFFIENYKQLDL 60  
+NHFFEL+KLKKGALTN NI NI++Y +K+ SLS+R+MAVVS CK+P+ FIE YKQLD+  
50 Sbjct: 1 VNHFELFKLKGALTNKNNILNILDY-QKHQKESLSLSDMAVVSCKRPSHFIRAYKQLDI 59  
Query: 61 KKLKQEFKKPPVLISILDENYPLELKEIYNPPVLLFYQGNIELLSKPKLAVVGARQASQIG 120  
+ L+ EFK+PP +SILD+YP+ LKEIYNPPVLLP+QGN++LL KPKLA+VG+R++S G  
60 Sbjct: 60 QNLKMEFKQFPSISILDKIYPMALKIETYNPPVLLFYQGNLLELKKPKLAIVGSRSSITG 119  
Query: 121 QCSVKKIKIETNNQFVIVSGLARGIDTAHVSAKNGGSSIAVIGSGLDVYYPTENKQLQ 180  
+SV+KI+KR N+FVIVSGLARGIDT+AH++ LKNGG +IA+IG+GLD +YP EN++LQ  
70 Sbjct: 120 VQSVRKILKELGNRFVIVSGLARGIDTSAHLAKLNGGQTIAIIGTGLDRFYPERKRELQ 179  
Query: 181 EYMSYNHLVLSEYFTGQPLKPHFPERNRIIAGLCOQIVVAEAKNRSGSLITCERALESG 240

-709-

```

 ++ NHLVL+EY GE+ L +HPPERNNIAGL +GI+V EAK RSGSLTIC+ +EBG
Sbjct: 180 TFLGHNHLVLTEYGPGEALSYHPPERNNIAGLSRGILVVEARNRSGSLTICQIGIEG 239

Query: 243 REVEAIFGNIIDKSDGCHHILQDGAKLISGKDILSEY 279
 R++FA+FGNI+DGKS+GC LI+EGA C+ SG DILSEY
Sbjct: 240 RDFAVFGNILDGKSGCLQIKGATCVTSGMDILSEY 278

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 619

A DNA sequence (GBSx0659) was identified in *S. agalactiae* <SEQ ID 1923> which encodes the amino acid sequence <SEQ ID 1924>. This protein is predicted to be lipoprotein (ceuE). Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAA06500 GB:AJ005352 lipoprotein [Staphylococcus aureus]
Identities = 122/348 (35%), Positives = 201/348 (57%), Gaps = 16/348 (4%)

Query: 1 MTKLLIALIALCTILITTSQAVLAKEKSO-----TVTIENNYSVYIKERKDKDPNK 52
 M K ++ +LA+ +L KE+ + TV I+NRY + + EK+ D K
Sbjct: 1 MKKTIVLVLAVMFLAAACGNNSDKESKSETKGSKDTVVIENNYFM--RGEKIDGSDAK 58

Query: 53 KOISETLKVPLKPKKVVVFDMZALDTITALGAESVIGIPKAKIALSLIPNVSVYKAK 112
 K + ET++VP P+ VV D GALT + +G V + PK + SL FN ++S +K
Sbjct: 59 K-VKETVEVPKPNENAVVLDYGALDVNKMGLSDKVKALPKGEGGKSL-PNFLES-FKDD 115

Query: 113 RYQDVSGLFEPNFEAIARMQPDVVLGARMASVNIIEKLEAAPKAALVYAGVDKSKVFD 172
 +Y +VG+L E NF+ IA +P+V+P+ R A+ N+++ K+AAPKA +YV G D K +
Sbjct: 116 KYTNVGNLKEVNFDKIAATKEPVI FISGRTANQJHLEDFKKAAPKAKIVVGADEKNLIG 175

Query: 173 KGVAERVTMLGKIPDQNGKATFNKDIQAQVLKLTQKTIKKGKPTALPVMANGSELITQS 232
 + + +GKI+D+ KAK NKD+ + + + K T +++ N GEL T
Sbjct: 176 S-MKQNTNIGIKYIDKEVKAELNKDLDNKIASMKDKTKNPNK-TVMYLVNREGELSTFG 233

Query: 233 PSGRPGW-IPSGVGPKAVNNEKLSHSTFPVSVEYIAEKNNYFLVLDRGATIGQGASSK 291
 P GRPG ++ GF AV++ S+HG VS EY++NPN++ +DRG + ++K
Sbjct: 234 PKGRPGGLVYDITLGPNAVDRKVSNSNGHQNVSNEYVKNKPNFVILMDRGQAVSGKSTAK 293

Query: 292 ELFNNVDIKATDAVNKKRVHVEDGDKWYINSGGSRVILRMKIDQVNF 339
 + NN V+K A+K +V+ +D K WY +G + T+ I++ V
Sbjct: 294 QALNNPVLKNVKAIEKDKVYNLDPKLWYFAAGSTTTTIKQIEELDKVY 341

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1925> which encodes the amino acid sequence <SEQ ID 1926>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-710-

An alignment of the GAS and GBS proteins is shown below:

Identities = 57/255 (22%), Positives = 104/255 (40%), Gaps = 30/255 (11%)

```

5 Query: 66 KKVVFVFMGALDTITAIAGKESVIGIPKAKNALSLPNNVSVVYKAKRYQVGSILRPNFP 125
 ++V + +D L + ++G +K L LP +V + VG P+
 Subject: 45 GRIVATSVAVVDICIRLALD--LVNVCSEK--LYTLPKRYIAVKR-----VGLFPNFDI 94

10 Query: 126 EALARMQHFUVFLGARMASVDNIEKLKERAPKAAALVYAGVDSKKVFKGVAKRVIMLGI 185
 E LA ++P + + E L+ K Y ++ + V +G+ + + LG +
 Subject: 95 ELIASLKPITWLSPNISIQ-----EDLEPKYQKLDTSYGFLNLRV--BOMYQSIDULGANL 147

15 Query: 186 FDQNKAKTTFKIDIAQAVLKLQKTIEKKGKPTALPUMNISELLTQSPSRFGPNIESVGS 245
 F + ++AK + Q + K KP L +H G L + G + + G
 Subject: 148 FQRQQRAKELAQYQDYTRAFQAKRKGKKPKVLLMLGSLVATNQSYVGNLLDLAG 207

20 Query: 246 FKAV---NENEKLSHSTPVSVEYIAEIGNPNYLFVLDRGATIQGAS---SKELFNIDVT 299
 + V +E E Ls++ E + K P+ +L I KE ND+
 Subject: 208 GENVYQSDKEKFLSANP-----EDMLAKEPD--LILRTAHALPDVKVMDFKFAENDIW 260

20 Query: 300 KATDAVGNKRVHVEVD 314
 K NVK ++V++D
 Subject: 261 KHPITAVKEKVIDLD 275

```

SEQ ID 1924 (GBS181) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 5; MW 38.7kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 3; MW 64kDa).

The GBS181-GST fusion product was purified (Figure 204, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 299), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 20

A DNA sequence (GBS00660) was identified in *S. agalactiae* <SEQ ID 1927> which encodes the amino acid sequence <SEQ ID 1928>. This protein is predicted to be iron(III) ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3231 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

45 >GP:CAB12190 GB:Z99106 similar to ferrichrome ABC transporter
 (ATP-binding protein) [Bacillus subtilis]
 Identities = 125/247 (50%), Positives = 187/247 (75%)

50 Query: 1 MIQINNHLKFPYQOKELIKDINISIPKGVTAALGPNSGKSTLLSCLSRLEPYDNGEIFL 60
 K+++ N+ K YG K +!++ ++I KGG+T+ +GPN+GKSTLLS +SRL D+GEI++
 Subject: 1 MVEVRNVSKQXGQKVLETSVITQKGTISFGPNAGKSTLLSIMSRLIKDQSGEIF 60

Query: 61 DKVFLARYSSNDIAKTLAILRQSNHLTIKIKVRDLIGFRFPFYSQKRLSQKDKAVIESVI 120
 D + S +LAK ++IL+Q+N + ++ ++DL+ FGRFPFYS+GRL++D I +
55 Subject: 61 DQGEIGACDSKELAKQWILKQANQINIRLITKDLVSGRFPFYSQKRLTSEEDVWHINQAL 120

```

-711-

Query: 121 SYNDINDIADEFINNLSGGQIQRAFIAMTMAQCTQYICLDBFIANNLDMKAVQNMIDLIKR 180  
 SYM L DI D+ + + LGGQ QRAFIAM +AQDT YI LDBFIANNLDMK+V+ +M L+KR  
 5 Subjct: 121 SYMKLSDIQCKYLDQLSGGQCORAFIAMVIAQDTYIFLDBFIANNLDMKHVSIEIMKLKR 180

Query: 181 YAYEFNKTIIVIIHDINFAHYADNVVALKKGQVVTGCTVEDVMGSKILSHLDFMPIRIE 240  
 E KTVI+IHDI+FA+ Y+D +VALK G+V G E+ + + +L + +DM I I+  
 10 Subjct: 181 LVEELGKTVIVIHDI+FAVSVDYI+VALKNGRIVKGGPPESVIETSVLEIYDITPIQ 240

Query: 241 TVDGKPI 247  
 T+D + I  
 10 Subjct: 241 TIDMQRI 247

There is also homology to SEQ ID 1930.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 621

A DNA sequence (GBSx0661) was identified in *S. agalactiae* <SEQ ID 1931> which encodes the amino acid sequence <SEQ ID 1932>. Analysis of this protein sequence reveals the following:

20 Possible site: 33  
 >>> Seems to have a cleavable N-term signal seq.

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -12.74 | Transmembrane | 271 - 287 ( 266 - 295) |
| INTEGRAL | Likelihood = -8.55  | Transmembrane | 49 - 65 ( 47 - 69)     |
| INTEGRAL | Likelihood = -8.07  | Transmembrane | 185 - 201 ( 178 - 207) |
| INTEGRAL | Likelihood = -7.70  | Transmembrane | 112 - 128 ( 105 - 132) |
| INTEGRAL | Likelihood = -7.38  | Transmembrane | 231 - 247 ( 227 - 261) |
| INTEGRAL | Likelihood = -2.50  | Transmembrane | 139 - 155 ( 135 - 156) |
| INTEGRAL | Likelihood = -1.97  | Transmembrane | 302 - 318 ( 301 - 319) |

25

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6095 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12189 GB:Z99106 similar to ferrichrome ABC transporter  
 (permease) [Bacillus subtilis]  
 Identities = 138/315 (43%), Positives = 222/315 (69%), Gaps = 6/315 (1%)

40 Query: 9 KLLILLILLIAAILFLFYGIPTDNRFLIYILKTRYQKLIALLVIGICIGSSLLIFQT 68  
 K+ +L+ L I I LFL Y + Y L R +K+ A+L G I S++IPQT  
 Subjct: 6 KIALLVLAIVICIGLFLFYDLGNWD-----YTLPRRIKKVAIVLTGGAIAFSTMIPQT 59

45 Query: 69 LTNKRLLTPSIIGLSLYILLIQGLMYLIGAQRVVKFSSSFSSFLSLLMVGFAYLFTFI 128  
 +TNKR+LTPSI+GLDSLY+LIQTG+++L G+ ++ + +F++S+LLM+ P+ +L+ I  
 Subjct: 60 LTNKRLLTPSIIGLSLYMLIQGLIIFLGSANMVIMNKNINFIISVLMILFSLVLYQI 119

50 Query: 129 LFRNKKQSLFYVLLAGLIFWTLFSSSISSFQAIMDNDPMILQNQLFASFWAINTKILMI 188  
 +F+ + +++P+LL G+ +F TLFSS+SSF+Q ++DN+ +F ++Q+++FASN INT +LM+  
 Subjct: 120 MFRGGRGNIFPLLIIIGVGTILFSSLSFPMLIDNRFQVQVQDMFASFNINFDLILMI 179

55 Query: 189 SFIIIVVSPVINWPFKEKLDVLLIGKENAISLISYQKLITRFFPLMLALMAVATALVGP 248  
 +FII +++ V W F K DVL LG+E+A++LGI Y K+ + + +A++V+++TALVGP  
 Subjct: 180 AFIIIFLLTGYYVWRPTEFFDVLISLGRHVAHLGDYDVVKQMLIVVAILVSVSTALVGP 239

60 Query: 249 ITFLGLLVARIITYRSHFFRHQILVPIAIVICIFTLVLGCHLVQNLHITVQLSVLLNLI 308  
 I FLGLLV ++ T+H L+ ++ I I LV GQ +V+ + + LSV+ +H  
 Subjct: 240 IMLFLGLLVNLAREFLKYKRSYLIAGSVFISIIALWGQFVVEKVFTEFTTSLSVINFA 299

Query: 309 GGSYFIPTLTKRKN 323  
 GG YFI+ L+K K+

-712-

Sbjct: 300 GGIIYFIYLLAKENKS 314

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1933> which encodes the amino acid sequence <SEQ ID 1934>. Analysis of this protein sequence reveals the following:

5      Possible site: 58  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL    Likelihood = -13.64    Transmembrane    33 - 49 ( 26 - 61)  
 INTEGRAL    Likelihood = -8.97    Transmembrane    259 - 275 ( 246 - 286)  
 10    INTEGRAL    Likelihood = -8.65    Transmembrane    296 - 312 ( 294 - 316)  
 INTEGRAL    Likelihood = -8.39    Transmembrane    83 - 99 ( 78 - 104)  
 INTEGRAL    Likelihood = -6.26    Transmembrane    212 - 228 ( 210 - 231)  
 INTEGRAL    Likelihood = -4.04    Transmembrane    113 - 129 ( 110 - 132)  
 INTEGRAL    Likelihood = -3.61    Transmembrane    140 - 156 ( 134 - 157)  
 INTEGRAL    Likelihood = -2.71    Transmembrane    165 - 181 ( 165 - 181)  
 15    INTEGRAL    Likelihood = -1.06    Transmembrane    327 - 343 ( 327 - 343)  
 INTEGRAL    Likelihood = -0.22    Transmembrane    50 - 66 ( 50 - 66)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6456(Affirmative) < succ>  
 20    bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9175> which encodes the amino acid sequence <SEQ ID 9176>. Analysis of this protein sequence reveals the following:

25      Possible site: 49  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL    Likelihood = -13.64    Transmembrane    24 - 40 ( 17 - 52)  
 INTEGRAL    Likelihood = -8.97    Transmembrane    250 - 266 ( 237 - 277)  
 30    INTEGRAL    Likelihood = -8.65    Transmembrane    287 - 303 ( 285 - 307)  
 INTEGRAL    Likelihood = -8.39    Transmembrane    74 - 90 ( 69 - 95)  
 INTEGRAL    Likelihood = -6.26    Transmembrane    203 - 219 ( 201 - 222)  
 INTEGRAL    Likelihood = -4.04    Transmembrane    104 - 120 ( 101 - 123)  
 INTEGRAL    Likelihood = -3.61    Transmembrane    131 - 147 ( 125 - 148)  
 INTEGRAL    Likelihood = -2.71    Transmembrane    156 - 172 ( 156 - 172)  
 35    INTEGRAL    Likelihood = -1.06    Transmembrane    318 - 334 ( 318 - 334)  
 INTEGRAL    Likelihood = -0.22    Transmembrane    41 - 57 ( 41 - 57)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.646(Affirmative) < succ>  
 40    bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 80/326 (24%), Positives = 157/326 (47%), Gaps = 34/326 (10%)  
 45    Query: 10 LLTLLTLLIRATILFLIYGIPTDANRFL-----IYILKTRFYKLIALLVIGICI 59  
 +L+L LL A+I + G+                    + I R+ +++ +L G I  
 Sbjct: 34 VLLLSLLFLAVIASLGLSGLVSGAIVGATVGLFVAYDQVALLYDLRPFRIIVALLAGAGI 93  
 50    Query: 60 GSSSLIFQILTNRLTPTSIIGL---DSLYKILIQTGLMYLIGAQRVIKSPSSSFL---L 113  
 S ++FQ + N + P+IIG+ S +L+ +L                    +++ + SFL +  
 Sbjct: 94 AVSGVLPQAVLKNPISDPAIGKICSGASFWLVSSLL-----PQLLYQPTVSPLOGGV 148  
 55    Query: 114 SLLLMGVFAYLLFTILFRKKQSLYFLVLAQLIWTLPSSSISSFTQIMDNPFMILNQ 173  
 S LL+ G A+                    K + ++L G+ N LF +S+ + + M+ N  
 Sbjct: 149 SFLLYGLAW-----KGLNPRLRLITGIALNALFMGLSTALTSFPTASPMV--NA 198  
 Query: 174 LAFSFAINTKI-LNISFTIIIVSVFNWPFPELKVLLGKRNATSLGISYQKLTTRFF 232  
 L A + T + + F + ++                    K +LLL + LGI L  
 60    Sbjct: 199 LLAGHSQRTWADVGVLPFTYTFIGLLALLSLKTCNLLLLDQVTRHGLDATALRLGVS 258  
 Query: 233 LMLALMVATATVAGVPTPLGLLVAHITHTSFTFRHQILVPIAIVICPTPLVQGLHQ 292  
 L L+ ++AT++VG ++PLGL+V H++                    +HQIL+P+ ++ F +L L +

-713-

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

Sbjct: 259 LVAVILASVATSI VGVSVFLGLIVPHMSRLLVGS-KHQIIPFSAIGAPVFLADTLGR 317

Query: 293 NILH-LTVGLSVLNLIGGSYPFTLL 317

+L + L + ++++++GG YFI+ L

Sbjct: 318 SLAYPLEISPAIMSVGGVPFTYLL 343

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2491> which encodes amino acid sequence <SEQ ID 2492>. An alignment of the GAS and GBS sequences follows:

Score = 51.9 bits (122), Expect = 5e-08  
 Identities = 73/327 (22%), Positives = 137/327 (41%), Gaps = 38/327 (11%)

Query: 494 IISSLGTAISTVAQSIGTGIAIAFRGLGAIAIMVPPITWLALGTAILMVGAAFAALGTQA 553  
 +I L T + G L IA +GA +V A+ L++ A

Sbjct: 573 VILGLVTTAAMLLGAIAPLVIAIGAIGAPVGVVAAIVGNAIVITLLIQAINMGA--- 629

Query: 554 DGISQLRTIGDXXXXXXXTDSLATLTLTIANAGSMPLIVAGAIQIVG-----A 606  
 I++ L++ D ++ T T A + ++G S +V +

Sbjct: 630 --ITELQSTWDSCHANKSELWNTNVTAT---TANSFTAMLSGLMSSVSVSTQGSWSS 684

Query: 607 VAGGLSOLIIVASTGVSLVIGAFITGLLGGI-SGVINSISAVIQSLTGVTIVPFGILATVI 665  
 LS + ++ TG + +PT L + SG+++ S + +L+ I+ +PNGI +

Sbjct: 685 FTSSLNIFPSLITGAQLMSSPFTSLNLSGLVSTGSLNPNLSSTISGIFNGILSTA 744

Query: 566 SSVGSTIKDVLITGLTAFEGFGNGVKSALGEGVAVIESFGSAVR-----NVLDGVAN 717  
 S++ ++IK ++ A+G N V N + GV A+ F ++ + G AN

Sbjct: 745 SNINWIKSTIS---NAIDGAVN N---GVNAIKNFPQIKWPHIPLHPRFVSGSAN 798

Query: 718 ILDSM--GTAALNAGRGVKEKARGIKMLVDLSGLDLVATLAASGLGNASSAGNMTLL 775  
 LD + G ++ G+ AKG ++ +L + A V G A +TL

Sbjct: 799 PLMLKGLPLI---GIDYKAG-GIMTKFTLPGMNGHRAMVGGEGAEALPLNKSTL 853

Query: 776 GSANSVKANGMTRLATSAITAITGLTV 802  
 G+ +AN M + + +G+T+

Sbjct: 854 GAIGQSILANTM-NTSNINNVNPSGVTI 879

Score = 33.2 bits (74), Expect = 0.019

Identities = 83/477 (17%), Positives = 175/477 (36%), Gaps = 103/477 (21%)

Query: 420 GSFLDKISTKFLGFLGKKAKETD-----QAANGSRKSGGIISQIFNGLNI 465  
 G + +++T+FGD G+K K ++ +A +++ LG +

Sbjct: 313 GDVAGELATQFGLTGKILKASSELLIKYAEINETDSSSAISAKQAIKAYLTAEGLMV 372

Query: 466 VKSAGTAISTAAGIGTGKIKTALSGAPPIISSLGTASTVA-----QGIGTGLAIA- 516  
 + + A + + T ++ A+ GAP I LG + A G+ + A++

Sbjct: 373 LDNVTKAAQDTGQVDITGLVTAAMLLGAIAPLVIAIGAIGAPVGVVAAIV 431

Query: 517 -----FRGLGAIAIMVPPIT--WLALGTAILMVGAAFAALGTQA----- 553  
 GL ++ + T AL A + G+ A A

Sbjct: 432 LSKAAVIYAKDGKLTLDGSLITVSAIQNSTSETALSIASEIFGSAAPKPMVDALGRGAF 491

Query: 554 --DGISQLRTIGDXXXXXXXTDSLATLTLT-----IANAGSMPLIVAGAIQIV 604  
 D +++ ++ D + L +A G +L V A+ ++

Sbjct: 492 SFDLAPAAKSSSGSTVSTTDETDLPIDIKITQVSNQAKGMAELGGKILSTVIPLELMV 551

Query: 605 GAVAGLS-----QLII--AVSTGVSLVIGAFITGL--LGGISGVINSISAVIQ 648  
 G + ++ Q I+ V+T V +++GA L +G I + + A I

Sbjct: 552 GNLESSVNNFTSLNETDQTTVILGLVTTAAMLLGAIAPLVIAIGAIGAPVGVVAAIV 611

Query: 649 SLTGVTIVAFNGI-----ATVISSVSGSTIKDVLITGLTAFEGFGNGVK 691  
 VIT + I A S + + I T + F + +G+

Sbjct: 612 GAIVATITLIQAINMNGAITEMLQSTWDSCHANKSELWNTNVTATANSFTAMLSGMA 671

Query: 692 SALEGVG-AVIESFGSAVRN V---LDGVANILDSMGTAAANAGRGVKEKARGIKMLVDL 746  
 S++ G++ SF S++ N+ + G + + S + N G+ +

Sbjct: 672 SSVVSTGQSLWSSPISLISNIFPSLITGAQLMSSPFTSLNLSGLVSTGSLN----- 725

Query: 747 SLGLDVATLAASGLGNASSAGNMTLLGSANSVKANGMTRLATSAITAITGLTV 803

-714-

+L +T++ + +G+ +++++ ++ S +S +G ++ AI L F  
 Sbjct: 726 -FNNLSSTLSGIPNGI--LSTASNIWNSIKSTISNAIDGAKKAVSNVNAIKNLNF 779

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 622

A DNA sequence (GBSx0662) was identified in *S. agalactiae* <SEQ ID 1935> which encodes the amino acid sequence <SEQ ID 1936>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial cytoplasm | --- | Certainty=0.2277(Affirmative) | < succ> |
| bacterial membrane  | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 623

A DNA sequence (GBSx0663) was identified in *S. agalactiae* <SEQ ID 1937> which encodes the amino acid sequence <SEQ ID 1938>. This protein is predicted to be membrane protein (ceuB). Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -11.30 | Transmembrane | 241 - 257 ( 237 - 274) |
| INTEGRAL | Likelihood = -6.42  | Transmembrane | 127 - 143 ( 118 - 149) |
| INTEGRAL | Likelihood = -5.79  | Transmembrane | 152 - 168 ( 150 - 174) |
| INTEGRAL | Likelihood = -5.47  | Transmembrane | 312 - 328 ( 309 - 330) |
| INTEGRAL | Likelihood = -4.83  | Transmembrane | 289 - 305 ( 287 - 308) |
| INTEGRAL | Likelihood = -4.67  | Transmembrane | 24 - 40 ( 22 - 46)     |
| INTEGRAL | Likelihood = -4.35  | Transmembrane | 69 - 85 ( 68 - 86)     |
| INTEGRAL | Likelihood = -4.19  | Transmembrane | 200 - 216 ( 198 - 216) |
| INTEGRAL | Likelihood = -2.76  | Transmembrane | 107 - 123 ( 107 - 123) |
| INTEGRAL | Likelihood = -0.85  | Transmembrane | 258 - 274 ( 258 - 274) |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5522(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

A related GBS nucleic acid sequence <SEQ ID 8621> which encodes amino acid sequence <SEQ ID 8622> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2  
 SRCFLG: 0  
 McG: Length of UR: 23  
 Peak Value of UR: 2.64  
 Net Charge of CR: 2

McG: Discrim Score: 8.59  
 GVH: Signal Score (-7.5): -4.6  
 Possible site: 26  
 >>> Seems to have an uncleavable N-term signal seq



-715-

Amino Acid Composition: calculated from 1  
 ALON program count: 9 value: -11.30 threshold: 0.0

|            |                     |               |                        |
|------------|---------------------|---------------|------------------------|
| INTEGRAL   | Likelihood = -11.30 | Transmembrane | 226 - 242 ( 222 - 259) |
| INTEGRAL   | Likelihood = -6.42  | Transmembrane | 112 - 128 ( 103 - 134) |
| INTEGRAL   | Likelihood = -5.79  | Transmembrane | 137 - 153 ( 135 - 159) |
| INTEGRAL   | Likelihood = -4.67  | Transmembrane | 9 - 25 ( 7 - 31)       |
| INTEGRAL   | Likelihood = -4.35  | Transmembrane | 54 - 70 ( 53 - 71)     |
| INTEGRAL   | Likelihood = -4.19  | Transmembrane | 185 - 201 ( 183 - 201) |
| INTEGRAL   | Likelihood = -3.08  | Transmembrane | 268 - 284 ( 265 - 284) |
| INTEGRAL   | Likelihood = -2.76  | Transmembrane | 92 - 108 ( 92 - 108)   |
| INTEGRAL   | Likelihood = -0.85  | Transmembrane | 243 - 259 ( 243 - 259) |
| PERIPHERAL | Likelihood = 5.73   | 203           |                        |

modified ALON score: 2.76  
 local HYDIP: 7 CFF: 0.552

\*\*\* Reasoning Step: 3

----- Final Results -----

|                                                              |
|--------------------------------------------------------------|
| bacterial membrane --- Certainty=0.5522(Affirmative) < succ> |
| bacterial outside --- Certainty=0.0000(Not Clear) < succ>    |
| bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>  |

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12188 GB:Z99106 similar to ferrichrome ABC transporter  
 (permease) [Bacillus subtilis]  
 Identities = 149/304 (49%), Positives = 234/304 (76%)

Query: 29 LVLISLSTSLFVGKESIPLQTHLDQSQVDIFLTSRLPRTISILISGASLSVCGLLMQQL 88  
 L+LL+TS+P+GV+ + + L + + SRLRP ISI+I+G S+S+CGL+MQQ+

Sbjct: 10 LILLAVTSVFIGVEDLSPLDLFSLKQFASLTPASRLPRLISIVLGLMSICGLIMQOI 69

Query: 89 TQKRFVSPTTSSTMWAKLGVVVTVLFFRNTSIFPIQLCIASGFAILGSLFVTLRMTF 148  
 +HKFVSPTT+GTMDW+LQ++++L+ F + S I++ +A FA+ G+ LF+ IL+ I F

Sbjct: 70 SKKRFVSPTTASSTMWAKLGLISLGLFPTSAPLHKMLVAFVFLAGNLFMKILRIKF 129

Query: 149 KNIIFIPILGLMLQIVAAATVFLSTHFQVLQSVNHWLQNFPSIMTSHRYEYLALPCL 208  
 D IPIL+GLMLG IV++ F+ + ++Q+V+SWLQ+FS++ RYE+LYL+P +

Sbjct: 130 MDTIPIPLVGLMLGNIVSISATFLAYKYDLIQVSSNLQDPSLVVKKRYEYLALPILN 189

Query: 209 FLVYFPAHQFTIVGLGESFAKNLGVAYEKMIVPGLNLVSIIMTSLVIIIVGALPFLGLIVP 268  
 + Y +A +FT+ G+GESF+ NLG+ Y++++ GL++VS++TSLVI+ VG LPFLGLI+P

Sbjct: 190 IIAVYADKFTLAGMGESFVNGLKCYKRVVNIIGLITVSLTSLVILVGMPLFLGLIIP 249

Query: 269 NLISITKGDHMSSTILETSLLGACTIVMICDLFGRIIVFPYEVSIGVTLVGLSAPFLISI 328  
 N++SI +GD++ S++ T LLGA V+ CD+ GR++IPFYE+SIG+ +G++GS FL +

Sbjct: 250 NIVSYRGDNLSKSLPHTVLGAVFVFLCDILGRITIIFFPYEISGLMVGITIGSGIFLPLML 309

Query: 329 IRNE 332  
 +R +

Sbjct: 310 LRRK 313

There is also homology to SEQ ID 1940.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 624

A DNA sequence (GBSx0664) was identified in *S. agalactiae* <SEQ ID 1941> which encodes the amino acid sequence <SEQ ID 1942>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -0.90 | Transmembrane | 140 - 156 ( 140 - 156) |
|----------|--------------------|---------------|------------------------|

-716-

----- Final Results -----

bacterial membrane --- Certainty=0.1362(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA006720 GB:AP001517 maltose transacetylase (maltose  
 O-acetyltransferase) [Bacillus halodurans]  
 Identities = 93/182 (51%), Positives = 125/182 (68%), Gaps = 2/182 (1%)  
 Query: 2 TEKEMLAGQYYRPSAPELKKDREVALKMMQAFNN--EDNSSKRNVLQKWFQATGKSIH 59  
 TEKEMLAG+ Y+ PEL KDRE A + + FN E +R ++++ FG+ G+S++  
 Sbjct: 3 TEKEMLAGERYKAWDPKLVDKDRARRLTFLNQTTETSEKQTELTIKLFGSMGESVN 62  
 Query: 60 MEQRFCVDYGCNTYGVGFYANFNQTFLDVCEIRIGDNOMFGPKCLLTPLHPLDPIERN 119  
 +E F CDYG NI+VG NF+ANF+ LDVCE+RIG NOM P + T HP+ P+ER  
 Sbjct: 63 IEFTFCVDYGYNIHVGNFANFDCVILDVCEVRIGANOMLAGVHIYATHTPIHPLERV 122  
 Query: 120 SLEYGAPITQIGNNVVLGGGVITLGGVVLGDMVVVVGAGSVVTKSFBNVVVIAGNPAKTIKIL 182  
 G EYG P+ I NNVW+GG + PGV +G+N V+ +GSVVTK NUV+AGNEAK+I+ +  
 Sbjct: 123 EGEYEGKPVITIRNNWIOGRAIVNPGVTIGNNVIAGSGSVVTKDVPENVVAGNEAKVITQT 184

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1943> which encodes the amino acid sequence <SEQ ID 1944>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4052(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/188 (36%), Positives = 101/188 (53%), Gaps = 13/188 (6%)  
 Query: 2 TEKEMLAGQYYRPSAPELKKDREVALKMMQAFN-----NEDNSSKRNVLQKWFQATG 53  
 TE +KM G+Y + D E+ K M A + +R+ +L + FG  
 Sbjct: 3 TEFDNTRGEWY-----DANFDSELTQKMMQGLCFDLNQLKPSREBERSAVINQLFG 57  
 Query: 54 TKSIRHMQRFVCDYGCNTYGVGFYANFNQTFLDVCEIRIGDNOMFGPKCLLTPLHPL 113  
 + + + + P+CDYG NI G+N +N N P+d +I +GDN GP+ T HPL  
 Sbjct: 58 SFEGVLILSPFICDYGNTYGVGFYANFNQTFLDVCEIRIGDNOMFGPKCLLTPLHPL 117  
 Query: 114 DPIERNSELEYGAPITQIGNNVVLGGGVITLGGVVLGDMVVVVGAGSVVTKSFBNVVVIAGN 173  
 D RN GLE ET IG+NW G V ++PGV +G +GSVVT N + AG  
 Sbjct: 118 DYKRRENGLEKALPITIGDMVWFGANVWVGVTIGSGCVIAGSGSVVTHDIPVNSLAGV 177  
 Query: 174 PAKTIKKL 181  
 P ++++K+  
 Sbjct: 178 PQGVVKRI 185

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 625

A DNA sequence (GBSx0665) was identified in *S. galactiae* <SEQ ID 1945> which encodes the amino acid sequence <SEQ ID 1946>. This protein is predicted to be ribonuclease H (mhB-2). Analysis of this protein sequence reveals the following:

Possible site: 32

-717-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 79 - 95 ( 79 - 95)

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 9823> which encodes amino acid sequence <SEQ ID 9824> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:CB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]

Identities = 128/249 (51%), Positives = 168/249 (67%)

Query: 4 TIKEIKALLETIVDLKDKRWQEQYQDSRAGVQKAILCRKNHQSDDLDEARLEQMAYEK 63

T+K+IK L+ + D +D + + D R VQ + Q K + + + H YE+

Sbjct: 5 TVKDIKRLQEVKDAQDPFIAQCENDPRKSVQTLVQWLKQAKKALKEQWVMYSYER 64

Query: 64 KLYIEHNLIAIGIDEVGRGSLAGPVVAANVILPNNCKIKHNSDKKIPKKHQRIYQNIL 123

+ LIAG+DEVGRGSLAGPVVA+AVILP C+I L DSKK+ +KK +E Y+ I+

Sbjct: 65 LARNKGFRLLIAGVDEVGRGSLAGPVVASAVILPEECILGLTDSKKLSEKKRETYELIM 124

Query: 124 DQRLAVGIGIQDSQCIDDINITYEATKHAMIDAVSHLSVAPENHLIDAMVLDLSIPOTKII 183

+ALAGIGI ++ ID+INITYEA+K AM+ A+ L5 P++LL+DAM L L Q II

Sbjct: 125 KEALAVGIGIVEATVIDEINITYEASMMAMVKAIQDLSDTFDYLLVDAMTLPLOTAQASII 184

Query: 184 KGDANSLIAAASIVAKVTRDKIMSDYDSTYPGYAFSPNAGYGTKEHLBGLQKYGTIPPH 243

KQDA S+SIDA + +AKVTRD+MS Y TYP Y F KN GYGTKEHL L YG T+ H

Sbjct: 185 KGDAKSVSIAAGACIAKVTRDRMSAYARTYPMYGFENKNGYGTKEHLAALAYGTPEHL 244

Query: 244 RKSFEPTKS 252

RK+P P+S

Sbjct: 245 RKTFAVQGS 253

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1947> which encodes the amino acid sequence <SEQ ID 1948>. Analysis of this protein sequence reveals the following:

Possible site: 50

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.53 Transmembrane 79 - 95 ( 79 - 95)

----- Final Results -----

bacterial membrane --- Certainty=0.1213(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the databases:

&gt;GP:CB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]

Identities = 130/252 (51%), Positives = 176/252 (69%), Gaps = 3/252 (1%)

Query: 4 SIKATKESLEAVTSILDLPFORLAYDTTSGVOKALSRQVIOELAREERLIRAMLSYEK 63

+K IK+ L+ V DP + D R VQ ++ K E A +E+ M SYE+

Sbjct: 5 TVKDIKRLQEVKDAQDPFIAQCENDPRKSVQTLVQWLKQAKKALKEQWVMYSYER 64

Query: 64 ALYKKGKATAGIDEVGRGSLAGPVVAACVILPKYCKIKHNSDKKIPKAKHRTIYQAVK 123

KG++ IAG+DEVGRGSLAGPVVA+ VILP+ C+I GL DSKK+ + K R Y+ +

Sbjct: 65 LARNKGFRLLIAGVDEVGRGSLAGPVVASAVILPEECILGLTDSKKLSEKKRETYELIM 124

Query: 124 RKALAGIGI IDNQLIDSVNIYENTKAMLEAIKQLBQOLTDFDYLLIDANITLDIAISQQ 183

++ALH+GIGI++ +ID+INITYEA+K AM++AI+ L FDYLL+DANITL + +Q

Sbjct: 125 KEALAVGIGIVEATVIDEINITYEASMMAMVKAIQDLS--DTPDYLLVDAMTLPLOTAQA 181

-718-

Query: 184 SIIKGDANSLIAAASIVAKVTRDQMMANYDRI FPGYDFARNAGYGTKEHLQGLKAYGIT 243  
 SI+KGDA S+SIAA + +AKVTRD+MM+ Y +P Y F KN GYTKHL+ L AYG T  
 Sbjct: 182 SIIKGDAXSVSIAAGACIAKVTDRMMMSAYASTYPMYGFKEHKGYGTKEHLEALAAYGPT 241

5 Query: 244 PIHRKSFEPVKS 255  
 +HRK+P PV+S  
 Sbjct: 242 ELHKKTFAPVQS 253

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 163/256 (63%), Positives = 203/256 (79%), Gaps = 3/256 (1%)

Query: 1 MMATIKKIKAILETIVDLKDKRQWEYQTDSSRAGVQKAILRKQNIQSDLDREARLEQMIV 60  
 M +IK IK LE + L D +QE TD+R+GVQKA+ R+K IQ+L RE RLE NL  
 Sbjct: 1 MPTSIKAIKBSLEAVTSLDPLFQELATTTRSGVQKALKSRQKVIQARLARESRLEAMIS 60

15 Query: 61 YERKLYIEHINLIAGIDEVGRGPLAGPVVAAAVILPPNCKIKHLNDSKKIPKKRHOBIYQ 120  
 YEK LY + IAGIDEVGRGPLAGPVVAA VILP CKIK LNSDKKIPK KH+ IYQ  
 Sbjct: 61 YEKALYKGYKAIAGIDEVGRGPLAGPVVAAACVILPKYCKIKGLNDSKKIPKAKHETIYQ 120

20 Query: 121 NILDQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLS---VAPEHLLIDMVLDSI 177  
 + ++ALA+GIGI D+Q ID+NIYEATK AM++A+ L P++LLIDAM LD++I  
 Sbjct: 121 AVKEKALAIIGIILNQILDSVNIYEATKAMLEAIKQLEGGQITQEDYLLIDAMTLOIAI 180

25 Query: 178 FQTKIILGDANSLIAAASIVAKVTRDKIMSDDYSTYGFYAFSNAGYGTKEHLGLQKY 237  
 Q I+KGDANSLIAAASIVAKVTRD++N++YD +FGY F+FNAGYGTKEHL+GL+ Y  
 Sbjct: 181 SQQSILKGDANSLIAAASIVAKVTRDQMMANYDRI FPGYDFARNAGYGTKEHLSGLKAY 240

Query: 238 GITPIHRKSFEPVKS 253  
 GITPIHRKSFEPVKS  
 30 Sbjct: 241 GITPIHRKSFEPVKS 256

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 626

35 A DNA sequence (GBSx0666) was identified in *S.agalactiae* <SEQ ID 1949> which encodes the amino acid sequence <SEQ ID 1950>. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1865(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 627

50 A DNA sequence (GBSx0667) was identified in *S.agalactiae* <SEQ ID 1951> which encodes the amino acid sequence <SEQ ID 1952>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

-719-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3034(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06195 GB:AP001515 unknown conserved protein [Bacillus halodurans]  
 Identities = 140/281 (49%), Positives = 195/281 (68%), Gaps = 5/281 (1%)

Query: 3 TIONFPFGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPMILTKIVGDKPKIMILNEAD 62  
 TQWFPFGEM-KARR+V E +K +D V L+DAR+PLSS+NPM+ +IV KP+L++LNLK D  
 Sbjct: 2 TIONFPFGHMAKARRVETKRLKLDVVIELDARVPLSSRNPMDEIVAHKPRVLVLNLKDD 61

Query: 63 LADPRTKRWDRFPYESQGLKTLAINSKQSTVKKVIDIAKILMSDKIANLRGRGIQKRTL 122  
 LADP +TKW F+E G L IN++ V +++ + L I R +G++ +  
 Sbjct: 62 LADPSKTKWTRFFEEGQATVLFINAGTGGVSRISPAQCTLAQALIEKQRAKMGKPRAI 121

Query: 123 RIMIIIGIPNAGKSTLMRLAGKKIAVGNKPGVTKGQNLKSNKELEILDTPGILNPKFK 182  
 R NI+GIPI GNKSTL+NRILA K+LA VG++PG+TK QQN+K KELE+LDTPGILNPKFK+  
 Sbjct: 122 RANIIIGIPNVGKSTLINRLASKRIAKVGRPGITKQQNIKVGKLELELDTPGILNPKFK 181

Query: 183 DELVGLKLAITGAIDKQLLPMDVETIFGILNYFKTYYPDRLKRFKSNLREDEAPEIIMAL 242  
 D+ G +LA TGAID+LL +V +F L Y + YPDRL +R+K L ++ + A+  
 Sbjct: 182 DQATGFLRLAATGAIDKQLLPMDVETIFGILNYFKTYYPDRLKRFKSNLREDEAPEIIMAL 241

Query: 243 TQKILGY-----KDDYDRFYNLFFVEVRDGLGRYTLDIVGE 278  
 +K G+ DTD+ + ++E+R G LGR TL++ G+  
 Sbjct: 242 GKXGHLISGGYIDYDKTAMILRELRAGTARITLKVPGK 282

A related DNA sequence was identified in *Spyogenes* <SEQ ID 1953> which encodes the amino acid sequence <SEQ ID 1954>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2688(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 247/282 (87%), Positives = 265/282 (93%)

Query: 1 MATIQWFPFGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPMILTKIVGDKPKIMILNK 60  
 MA IQWFPFGHMSKARRQVQEN+KHVDFVTILVDARLPLSSQNPMILTKIVGDKPKIMILNK  
 Sbjct: 1 MATIQWFPFGHMSKARRQVQENVKHVDFTVLVDARLPLSSQNPMILTKIVGDKPKIMILNK 60

Query: 61 ADLADPRTKRWDRFPYESQGLKTLAINSKQSTVKKVIDIAKILMSDKIANLRGRGIQKE 120  
 ADLAD RTKRW+ +YESQ+KTLAINSKQSTVKKVPT+ AK IM+DKT LR RGIGKE  
 Sbjct: 61 ADLADATRTKRWKAYYESQGIKTLAINSKQSTVKKVTEANKELNADKIQLRGRGIQKE 120

Query: 121 TLRMTIIGIPNAGKSTLMRLAGKKIAVGNKPGVTKGQNLKSNKELEILDTPGILNPKFK 180  
 TLRMTIIGIPNAGKSTLMRLAGKKIAVGNKPGVTKGQNLKSNKELEILDTPGILNPKFK  
 Sbjct: 121 TLRMTIIGIPNAGKSTLMRLAGKKIAVGNKPGVTKGQNLKSNKELEILDTPGILNPKFK 180

Query: 181 FEDELVGLKLAITGAIDKQLLPMDVETIFGILNYFKTYYPDRLKRFKSNLREDEAPEIIM 240  
 FEDELVGLKLAITGAIDKQLLPMDVETIFGILNYF+ YYP+RL +RKF+ L+E+APEIIM  
 Sbjct: 181 FEDELVGLKLAITGAIDKQLLPMDVETIFGILNYFREYIPRILTKRFNIPLEBEAPEIIM 240

Query: 241 ALTKILGYRDDYDRFYNLFFVEVRDGLGRYTLDIVGEHDGN 282  
 LT++LG+DDYDRFY LPVEVRDGLG+YTL DIVG+ D+  
 Sbjct: 241 TLTRQLGFKDDYDRFYTLPVEVRDGLGQYTLDIVGQMDAD 282

-720-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 628

A DNA sequence (GBSx0668) was identified in *S.agalactiae* <SEQ ID 1955> which encodes the amino acid sequence <SEQ ID 1956>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9825> which encodes amino acid sequence <SEQ ID 9826> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CB12129 GB:Z99105 similar to hypothetical proteins [Bacillus subtilis]
Identities = 69/173 (39%), Positives = 102/173 (58%), Gaps = 13/173 (7%)

Query: 29 DKAKEKASV-----IKQASQTSQTSKKEVLQKKT---YPLNKEYENLEIHSVSTQTM 79
D A+E AS+ ++ + +T+K + K YP+ K ++ I V+ Q
Sbjct: 22 DHAEBHASINTKKTVENITDVRKTAKTSIDMTKPSGGSEYFDI-KQKHVWIDVNVKEQKAY 80

Query: 80 ITENDKVIPIKTIIVSTG---AKESPTPKGTFTVISPERGDPFFYNASSKEGAYYWSFKENGI 136
I I+ ++S+G K+ TPKGTF +SPERG+P++ +EGA YWVS+K HG
Sbjct: 81 IKEGNTITYTMMISSGLDQTKDDATPKGTFTVYEPERGERWFFSEGYQEGASYWWSKKNHGE 140

Query: 137 YLFHSVPTDQGNIEPEAKQLGKAASHGCVMSRADAIGFYENIPQGTITVTI 169
+LFHSVP + I EA++LG SHGC+R++ DAKW YENIP+ T V I
Sbjct: 141 FLFHSVFMTDQGVIKTSAEKLGTGVSHGCIRLITPDAGVYENIPFHTKVV 193
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1956 (GBS644) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 2 & 3; MW 49.6kDa) and in Figure 186 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 5-7; MW 24.6kDa) and in Figure 177 (lane 3; MW 25kDa).

GBS644-GST was purified as shown in Figure 236, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 629

A DNA sequence (GBSx0669) was identified in *S.agalactiae* <SEQ ID 1957> which encodes the amino acid sequence <SEQ ID 1958>. This protein is predicted to be carbon starvation protein A. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL likelihood = -11.25 Transmembrane 129 - 145 (122 - 157)
INTEGRAL likelihood = -9.92 Transmembrane 316 - 332 (305 - 342)
INTEGRAL likelihood = -6.42 Transmembrane 164 - 180 (157 - 181)
INTEGRAL likelihood = -5.73 Transmembrane 443 - 459 (441 - 466)
```

-721-

INTEGRAL Likelihood = -5.57 Transmembrane 416 - 432 ( 414 - 435)  
 INTEGRAL Likelihood = -4.88 Transmembrane 190 - 206 ( 183 - 209)  
 INTEGRAL Likelihood = -4.83 Transmembrane 78 - 94 ( 70 - 95)  
 5 INTEGRAL Likelihood = -3.13 Transmembrane 362 - 378 ( 359 - 379)  
 INTEGRAL Likelihood = -2.34 Transmembrane 228 - 244 ( 227 - 245)  
 INTEGRAL Likelihood = -2.02 Transmembrane 2 - 18 ( 1 - 18)  
 INTEGRAL Likelihood = -1.28 Transmembrane 393 - 409 ( 393 - 410)

----- Final Results -----  
 10 bacterial membrane --- Certainty=0.5501(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:AA793852 GB:AB004154 carbon starvation protein A, putative  
 [Vibrio cholerae]  
 Identities = 220/470 (46%), Positives = 311/470 (65%), Gaps = 16/470 (3%)

20 Query: 1 MVTFLGQVALLIVGYPTTYGRYIEKNFQIDENRQTAPBALADGYDFVPMFQWGMIELLN 60  
 M+ FL VA L+ GYF YG ++EK F I+E RQTPA DG D+VPM K +++LLN  
 Sbjet: 1 MLVFLTCAVALVGQYFIYGAFAVEKVFQINEKRQTPAHKTKDGVDFVPMSTFKVYLQLLN 60

25 Query: 61 TACTGPITGPILGALYGFVAYIWIIVLCIFAGAVHDVIMQISLRNNGALPELASRYLG 120  
 TAG GPITGPIT+GALYGF A +WIV+GCTFAGAVHDY GM+S+RN GA +P + RYLG  
 Sbjet: 61 TAGVGPITGPITMGALYGPAAWMLWIVGCTFAGAVHDYSGMLSIRNGGASVPTITRYLG 120

30 Query: 121 KSMKHVINIFSMILLILVATVFFVTPANLILSLPAG--TLGLPWIGLIFVYLISTV 177  
 KH +NIF+++LL+LV VVF PA +I +++ T+S+ ++ +IF YI+YI+T+  
 Sbjet: 121 NGAKHFMNIFAIIVLLLVGVFVSAPAGMTNLLNQDFTVSMITMVIIYAIYI+LATI 180

35 Query: 178 LPIDNALGRVYVPV-----CVILWSTRAVGLLITGQFDMPLNTFETFGNHPAGLG 230  
 +PADK +GA YP+ V LM + A + GGF++ ++ RN+P+ +  
 Sbjet: 181 VPDVGLIIGRFYPLFGALLIMVGLMTAIFSSHQVLOGFISDMV---KILATNDMP 236

40 Query: 231 IFPALFPTTISQAGISGFHATQAPVBRITVNERBGRPTTYGMIIABGVMIAGAGASNL 290  
 ++PALF TI+CAISGFHATQ+P+NR NR+ GRF FYQ NI EG+IA+IW ++S  
 Sbjet: 237 LNPALFITTACAGISGFHATQSPIMARCMEISQWRFFVYGAMIGEGIIALWCTVALSF 296

45 Query: 291 FKQ-QNLVEMTAACTPSAVVNVQVLMMLLSGVIGTIALIGIVTLVPSSGLGAPSLRTIVA 349  
 F + L E + G P VV LLG G IA +GV+LP+SG +AFES R IA+  
 Sbjet: 297 FGSLEALSEAVKNGGPGWVVGASFGQLGVFGVIAFLGVVITPISGDTAFRSRLIA 356

50 Query: 350 DYIHWKDTLPKIFAVTITPLVVISFVLTHVDFNLLWYENWANOQVAVIGLWATKYLIL 409  
 +Y +++Q TL + +PL+VI VLT VDF ++WEVF +ANQ TAV+ L AT XL+  
 Sbjet: 357 EYFNMQKTLRNKLLMAVFLVIGAVLTQVDFGIINWYFGFANQATVWMLTATATYLMR 416

Query: 410 KRNRYWVTVEARMFMILYAVVVIL-SQPIGENMLGILTYSLALVLTGIL 458  
 + +W+ VEA+PM + +IL S +GF + + I T + L G L  
 Sbjet: 417 HNKLEWITCTVEALFWTTVCISFILNSSLTGLPGLPMQISTAGVLAELGAL 466

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8623> and protein <SEQ ID 8624> were also identified. Analysis of this protein sequence reveals the following:

15 Lipop: Possible site: -1 Cread: 0  
 MoG: Discrim Score: 6.07  
 GvH: Signal Score (-7.5): -3.54  
 Possible site: 19

20 >>> Seems to have an uncleavable N-term signal seq  
 ALCO program count: 11 value: -11.25 threshold: 0.0

60 INTEGRAL Likelihood = -11.25 Transmembrane 129 - 145 ( 122 - 157)  
 INTEGRAL Likelihood = -9.92 Transmembrane 316 - 332 ( 305 - 342)  
 INTEGRAL Likelihood = -6.42 Transmembrane 164 - 180 ( 157 - 181)  
 INTEGRAL Likelihood = -5.57 Transmembrane 416 - 432 ( 414 - 435)  
 INTEGRAL Likelihood = -4.88 Transmembrane 190 - 206 ( 183 - 209)

-722-

```

INTEGRAL Likelihood = -4.83 Transmembrane 78 - 94 (70 - 95)
INTEGRAL Likelihood = -4.67 Transmembrane 445 - 461 (441 - 463)
INTEGRAL Likelihood = -3.13 Transmembrane 362 - 378 (359 - 379)
INTEGRAL Likelihood = -2.34 Transmembrane 228 - 244 (227 - 245)
5 INTEGRAL Likelihood = -2.02 Transmembrane 2 - 18 (1 - 18)
INTEGRAL Likelihood = -1.28 Transmembrane 393 - 409 (393 - 410)
PERIPHERAL Likelihood = 0.21 272
modified ALQW score: 2.75

10 *** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

20 ORF01729(301 - 1668 of 2082)  
 GP|9655126[gb|AAFY93852.1|AE004154(1 - 464 of 494) carbon starvation protein A, putative  
 (Vibrio cholerae)  
 %Match = 29.9  
 %Identity = 47.6 %Similarity = 68.6  
 Matches = 218 Mismatches = 138 Conservative Sub.s = 96

25            174            204            234            264            294            324            354            384  
TNEKLFIKRLIFISKQPFILKIGNFNFMILY\*SHENA\*\*N\*AKKFLGSGDMVTLGGVALLIIVGYFTYGRVYEKNFQI  
[: :| | | | | | | | | | :| :| | |  
MLWFLTCAALVGGFYIYGAFVEKVFGI  
               10            20

414            444            474            504            534            564            594            624  
DENRQTPAEALRDGYDFVPMWNGNIEILLIAGTGPIRGPIGLGALYGFAYIWIIVLGCIFAGAVHDYMIIGMISLRNG  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |  
NEKRQTPAHKTDGDVDYFMSTPKVYLQLLIAGVGPIFGPIGLGALYGPAMLMWVGVCIFAGAVEDYSGSMLIRNGS  
35            40            50            60            70            80            90            100

40

|                                                                                                     |     |     |     |     |     |     |     |
|-----------------------------------------------------------------------------------------------------|-----|-----|-----|-----|-----|-----|-----|
| 654                                                                                                 | 684 | 714 | 744 | 774 | 795 | 825 | 855 |
| ATLPELASRSLKGVGVINIPSMILLILVATVPVPTANLILSILPAGT---LSEFWIIGLIFYVYLISITVLVPIDKALG                     |     |     |     |     |     |     |     |
| :   :         :     :     :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : |     |     |     |     |     |     |     |
| ASVPSPTSLRVLGNAGKHPMNI FAIVLVLGGVGVFSAPAGMTNLINHQDIDPTVSMITMVVIFYLATILYFATVLPVDKILG                 |     |     |     |     |     |     |     |
| 120                                                                                                 |     | 140 |     | 160 |     | 180 |     |

[illegible][illegible][illegible]

65

|  |                                                                             |      |      |      |      |      |      |      |
|--|-----------------------------------------------------------------------------|------|------|------|------|------|------|------|
|  | 1581                                                                        | 1608 | 1638 | 1668 | 1698 | 1728 | 1758 | 1788 |
|  | FVPAFMFLYVVYYIL-SQPIPMNMLGILTYSIALVLVTGXVGLFWKSGQKKLTVHPEAFLFNDRHPINNYSSLS* | Y    |      |      |      |      |      |      |
|  | [ :] :] : :  [ :   : :   :   :                                              |      |      |      |      |      |      |      |
|  | TVPALFMTVCISFILNSTSLGFLPMDISTAGVLASLGLAYVAVKVSQKGEITDLAREKPOGVTKTA          |      |      |      |      |      |      |      |
|  | 440                                                                         | 450  | 460  | 470  | 480  | 490  |      |      |



Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 630

A DNA sequence (GBSx0670) was identified in *S. galactiae* <SEQ ID 1959> which encodes the amino acid sequence <SEQ ID 1960>. This protein is predicted to be lytR (lytR). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.80 Transmembrane 27 - 43 (27 - 43)

----- Final Results -----
bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP.ARB48183 GB:L42945 lytR [Staphylococcus aureus]
Identities = 93/245 (37%), Positives = 150/245 (60%), Gaps = 3/245 (1%)

Query: 1 MKVLVVDDEFVARNELIYLLNKYDNLVIAEAHDMATALAILLETFTDVALDIDHLRIDS 60
MK L++DDEP+ARNEL YLLN+ I EA ++ L LL +D+ LD++L D++
Sbjct: 1 MKALI IDDEPLARNELIYLLNKIGGFEEINRAENVKSTLEALLINCYDIIFLDVNLMDEN 60

Query: 61 GLQLAEYINRMPKPLLIIFATAYDQVAIQAFEHARDYLLKPYDFDLKQMDRVKGLS 120
G++L I RM +PP +IFATA+DOXA+QAFE +A DY+LKP+ R++QA+++V+ +
Sbjct: 61 GIELGAKIQKQKPPAILIFATAHDQAVQAPELNATDYLLKPPQOKRIEQAVNKVRATKA 120

Query: 121 TSTIIESVTSGLP--FKQYPLTVEDRIYLVASDDILLIEAMQKLIITPDKMYEIDG 177
S + + F Q P+ ++D+I+++ +I+ I G I T + YE
Sbjct: 121 KDDNNSAIAINMDSANFDQSLVVEIDDKIHMLKQCNIIIGIOTHNGITTIHTNHKYETTE 180

Query: 178 SLQQWQDKLPSSQFVVRHSYIVMNAIKTIEPWFMTQLQLHLNCKITVPVSRANVKPLK 237
L +++ -L + F+R+HRSYI+N IK ++ WFN T + L N + + V R+ + K K
Sbjct: 181 PLNRYEKRLNFTYFIRIHSYIINTYKIKVQQWPNYTYTMVLITNGVGMQVRSFMDPK 240

Query: 238 QMLGI 242
+G+
Sbjct: 241 ASIGL 245
```

There is also homology to SEQ ID 460.

SEQ ID 1960 (GBS399) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 7; MW 30.4kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 2; MW 55kDa). Purified GBS399-GST is shown in Figure 217, lane 9; purified GBS399d-GST is shown in Figure 236, lane 3.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 631

A DNA sequence (GBSx0671) was identified in *S. galactiae* <SEQ ID 1961> which encodes the amino acid sequence <SEQ ID 1962>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -7.59 Transmembrane 95 - 111 (86 - 116)
```

-724-

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -5.95 | Transmembrane | 155 - 171 ( 152 - 176) |
| INTEGRAL | Likelihood = -2.28 | Transmembrane | 189 - 205 ( 187 - 206) |
| INTEGRAL | Likelihood = -1.49 | Transmembrane | 122 - 138 ( 121 - 138) |

## 5 ----- Final Results -----

bacterial membrane --- Certainty=0.4036 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

## 10 The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:ABA48182 GB:L42945 lytS [Staphylococcus aureus]

Identities = 264/570 (46%), Positives = 389/570 (67%), Gaps = 2/570 (0%)

15 Query: 1 MTLFLDMERAGLIIILAYAFVHIPPIKQTLKQELAKKHQVILLILFSLFAIISNTGVE 60  
 ++L +++ER GLII+LAY ++IP+ K + + K ++ L I+PSLFA++SN TG+  
 Sbjct: 2 LSLTLALLERVLIIILAYVLMNIPYFKMLNNRRKTNKRWQLCIIFSLFALMSNLGITIV 61

20 Query: 61 IQSDLSIIPQTLNHIADQSSVANTRVLITGVSGLIGGPIVGIIVGLISVFRVYLGQGLAP 120  
 I S+ + D S+ANTRVLITGV+GL+GGP VG+ VG++S R GG  
 Sbjct: 62 IDHQSLSGSVYFRLLDDVSLANTRVLITGVAGLVGGPFVGLFVGVISGIFRVTMGAD 121

25 Query: 121 HIYVSSILLIGLCSGLSGNYLRKYNKIRVLDAMVVGFGMILQMCILIFSVDFNQALK 180  
 +Y+ISS+ IG+ +G G ++ + + +G ME++QM+ IL FS D A+  
 Sbjct: 122 QVYLISIFIGIIGYFGLQAQRKRYPISAKSAMIGIVMBMIQMSILTFSHOKAYAD 181

30 Query: 181 LVFSFISPMILSNTLGLGIFISISSTQKLESHAKAPQTHQVLELANITLPLRKLHTE 240  
 L+S I++PMI+ N++G IP+SII T K E+ K QTH VL+L N T FY ++GL E  
 Sbjct: 182 LISLIALPMIIVNSVGPALFMSIIITPKQGDQMKPVQTHDVLQMNITFPYFKELARE 241

35 Query: 241 SCQPVARIHKNMDVASLTSQSAIAYFGDGADHPLPWQILITKLAKRAIDNKRVSVA 300  
 S Q ++ II N VS+V++TS++ IL++VG G+DHH+P +ILT L+K + +OK+  
 Sbjct: 242 SAQQTAMIKNMLKVSVAITSKNELKSHVGGSDHIFITNELITSLSKDVLKSGKLV 301

40 Query: 301 TKDSIECDHKNCLPSIAIVPLHNTDVTGTLKPSDAGMHTVDRQLASGLNIPST 360  
 K EI C H NCL +ATVPL +H IVTDLK+YE++ +T+V+GLASGL HTSA+  
 Sbjct: 302 HTKEETGCSHPNCLPRAATVPLSMHSGIVGTLMWTFINPDLTFVVERQLASGLNIPSS 361

45 Query: 361 QLALQAEATRLIQLDAEMKSLQAVNHPPLFNAINITYGLRMDSEKARKLVQDFSKVI 420  
 Q+ LG+AE ++LL+DAE+KSLQAV+PHF FN++N I L+R++SEKAR+L+ + S  
 Sbjct: 362 QTELGEATQSKLLKDAEIKSLQAVSHPHFFNSINPISALVRINSEKARELLLSYFF 421

50 Query: 421 RANLQRAKNILPLHDELQUNAYLAEEARFPNMFNPLNQLNINSDNLMIPPPTLQVL 480  
 RANLQ +KQ+ I L EL QV AYL+AR+P N++ + D +++PFF +Q+L  
 Sbjct: 422 RANLQSSQHTITLDELQSVRAYLSLQARYPGRFNININVEDRYRD-VLVPFPLIQL 480

55 Query: 481 IENSYKHFAPKHNKQWLVKVTIARNN-DRLHIVQONGIGPKKELITLGGKTQISEQS 539  
 +EN+ KDAF + + N + V++ + + IIVQNG GI K+K+ LG+ + S+ G+  
 Sbjct: 481 VENAIGHAFNRRKQNDIDVSIKETHYHVRITVQNGQGISKDMELLEGTSVESSEGT 540

50 Query: 540 GTALENVLRLNIIVDQASLKFESNDSGT 569  
 G+A+ENL RL ++ A+L+FES SGT  
 Sbjct: 541 GSALENINLRKLKGLFGKRALQFESTSSGT 570

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1963> which encodes the amino acid  
 55 sequence <SEQ ID 1964>. Analysis of this protein sequence reveals the following:

Possible site: 39

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -6.79 | Transmembrane | 283 - 299 ( 276 - 307) |
| INTEGRAL | Likelihood = -5.57 | Transmembrane | 27 - 43 ( 24 - 48)     |

## 65 ----- Final Results -----

bacterial membrane --- Certainty=0.3718 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```

>GP:CAH54576 GB:AJ006396 histidine kinase [Streptococcus pneumoniae]
Identities = 115/231 (49%), Positives = 159/231 (68%), Gaps = 7/231 (3%)

5 Query: 351 MLASIKAYIDEVYVLEVBORDAQMRALQSOINPHFLYNTLEYIRMYALSQOQELADVIY 410
 ML ++ I ++Y LE+ Q+DA MRALQ-QINPHF+YNTLE++RMYA+ Q+ELAD+IY
 Sbjct: 1 MLDRLKINIHDYIQLELSQKDANMRALQAOINPHFYNTLEFLRMYAVMQSQDELADIY 60

10 Query: 411 AFASLLERNISQDKMTTLKEELAFCEKYIYLYQMYDFSFAYHKVIDESVADLAIPKFVI 470
 F+SLLENNIS ++ T LK+EL FC KY YL +RYP S AY KID ++ ++ IPKF +
 Sbjct: 61 EFSSLLERNISDERETLLQQLAEFCRKYSYLOMRYEFSIAYGFKIDPELENNKIKIPKFTL 120

15 Query: 471 QPLVENYFVHGIDYSRHDNALSIKALDETHLLIQVLNDNGRISQERLADMEKRLQ---- 526
 QPLVENYF HG+D+ R DN +SIKAL + + I V+DNKRG+S E+LA++ ++L
 Sbjct: 121 QPLVENYFAHGVDHRRRTNVISIKALKQDGFVEILVVDNNGRMSAEKLANIRKLSQRYF 180

 Query: 527 EQHQT---GNSSIGLQNVYLRIPHFFRDRVSWSMKEPNNGGFIQIRIRKD 574
 EQH + SIG+ HV+ R +F DR + ++ G +I I+ +
 Sbjct: 181 EQHQSYSQDQSGIIVNHERFVLYGDRYALTIESAQAGVQYRITTIQDE 231

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 59/180 (32%), Positives = 97/180 (53%), Gaps = 8/180 (4%)

25 Query: 375 QDAEMKSLQAVNPFLPALNTI--YGLIRMDSEKARKLVQDFSKVIRANLQRAKNLI 432
 +DA++++LQ+Q+NPFL+N L I Y L E A ++ F+ ++R N+ + K +
 Sbjct: 370 RDAQMRALQSOINPHFLYNTLEYIRMYALSQOQELA-DVIYAFASLLRNISQDK--MT 426

 Query: 433 PLHDSLEQWYALAEERFPNMFVFNLDNQTNDSNDLMIPFPTLQVLNIENSYKHAFIGV 492
 L +EL Y+ L + R+P+ A+++ + D L IP F +Q L+EN + H +
 Sbjct: 427 TLKEELAFCEKYIYLYQMYDFSFAYHKVIDESVAD-LAIPKFVIQPLVENYFVHGIDYS 485

 Query: 493 NKNQKVTIARNRDLHIIVQDNGIGIPKEKLITLQKTKQISQK--GSQTAIENLVRR 550
 +N L + D L I V DNG GI +E+L + K+ Q + S ++N+ RL
 Sbjct: 486 RHDNALSIKALDETHLLIQVLNDNGRISQERLADMEKRLQEHQHTGNSSIGLQNVYLR 545

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 632

40 A DNA sequence (GBS0672) was identified in *S. agalactiae* <SEQ ID 1965> which encodes the amino acid sequence <SEQ ID 1966>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> May be a lipoprotein

```

```

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

50 A related GBS nucleic acid sequence <SEQ ID 9827> which encodes amino acid sequence <SEQ ID 9828> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 633**

A DNA sequence (GBSx0673) was identified in *S.agalactiae* <SEQ ID 1967> which encodes the amino acid sequence <SEQ ID 1968>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -9.55 Transmembrane 52 - 68 (45 - 74)
INTEGRAL Likelihood = -9.18 Transmembrane 83 - 99 (76 - 106)
INTEGRAL Likelihood = -8.76 Transmembrane 126 - 142 (118 - 146)
INTEGRAL Likelihood = -7.48 Transmembrane 174 - 190 (170 - 191)
INTEGRAL Likelihood = -3.66 Transmembrane 195 - 211 (193 - 212)
INTEGRAL Likelihood = -1.28 Transmembrane 24 - 40 (24 - 40)

----- Final Results -----
bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8625> and protein <SEQ ID 8626> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: -8.54
GVH: Signal Score (-7.5): -5.6
Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program count: 6 value: -9.55 threshold: 0.0
INTEGRAL Likelihood = -9.55 Transmembrane 52 - 68 (45 - 74)
INTEGRAL Likelihood = -9.18 Transmembrane 83 - 99 (76 - 106)
INTEGRAL Likelihood = -8.76 Transmembrane 126 - 142 (118 - 146)
INTEGRAL Likelihood = -7.48 Transmembrane 174 - 190 (170 - 191)
INTEGRAL Likelihood = -3.66 Transmembrane 195 - 211 (193 - 212)
INTEGRAL Likelihood = -1.28 Transmembrane 24 - 40 (24 - 40)
PERIPHERAL Likelihood = 13.05 100
modified ALOM score: 2.41

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 634**

A DNA sequence (GBSx0674) was identified in *S.agalactiae* <SEQ ID 1969> which encodes the amino acid sequence <SEQ ID 1970>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.53 Transmembrane 83 - 99 (83 - 99)

----- Final Results -----
bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-727-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 635

A DNA sequence (GBSx0675) was identified in *S.agalactiae* <SEQ ID 1971> which encodes the amino acid sequence <SEQ ID 1972>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1902 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 636

A DNA sequence (GBSx0676) was identified in *S.agalactiae* <SEQ ID 1973> which encodes the amino acid sequence <SEQ ID 1974>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4763 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 637

A DNA sequence (GBSx0677) was identified in *S.agalactiae* <SEQ ID 1975> which encodes the amino acid sequence <SEQ ID 1976>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5089 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 5 Example 638

A DNA sequence (GBSx0678) was identified in *S.agalactiae* <SEQ ID 1977> which encodes the amino acid sequence <SEQ ID 1978>. Analysis of this protein sequence reveals the following:

Possible site: 25  
>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1978 (GBS184) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 7; MW 21kDa), in Figure 168 (lane 14-16; MW 36kDa – thioredoxin fusion) and in Figure 238 (lane 9; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 7; MW 46.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 639

A DNA sequence (GBSx0679) was identified in *S.agalactiae* <SEQ ID 1979> which encodes the amino acid sequence <SEQ ID 1980>. Analysis of this protein sequence reveals the following:

Possible site: 52  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 640

A DNA sequence (GBSx0680) was identified in *S.agalactiae* <SEQ ID 1981> which encodes the amino acid sequence <SEQ ID 1982>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

-729-

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2166 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9351> which encodes amino acid sequence <SEQ ID 9352> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1983> which encodes the amino acid sequence <SEQ ID 1984>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -3.77 Transmembrane 9 - 25 ( 5 - 27 )  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2508 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 64/86 (74%), Positives = 76/86 (87%)  
 Query: 1 MNGGGEWKNKPGYQITHEAKTYAISFSFGAGADRTYGHVAIVEDVKEDGSIPISESIV 60  
 MNGGGEW+ KPG+ TTH+ K GY +SF+PGQAGAD TYGHVA+VE +KEDGSI ISESIV  
 Sbjct: 452 MNGGGEWQRKPGFVTHKPFVGVVVSFAPGQAGADATYGHVAVEIGKEDGSILLSESIV 511  
 Query: 61 IGLGTISYRTFSFAAQAQLTVVVGK 86  
 +GLGTISYRTF+A +A+ LTVVVG+K  
 Sbjct: 512 NGLGTISYRTFTARQASLLTVVVGDK 537

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 641

A DNA sequence (GBSx0681) was identified in *S.agalactiae* <SEQ ID 1985> which encodes the amino acid sequence <SEQ ID 1986>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2495 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

AAB52379 GB:U31811 immunogenic secreted protein precursor [Streptococcus pyogenes]  
 Identities = 133/259 (51%), Positives = 170/259 (65%), Gaps = 4/259 (1%)  
 Query: 3 PSQPVTAIPQKSEVVTFAITSGDLPDVAIPYMASAAYVKNHIGNDAYTHNILLSRYG 62  
 P QP + A + V P S DL + P++ +SAAYV+HW G+ AYTHNILLS RYG  
 Sbjct: 174 PIQPFLLGA---APVPAPWRSSDKDLKLLK-PSRSEAAVVRHWITGDSAYTHNILLSRYG 229  
 Query: 63 ITAAQLDGLFLSTGTIVYDSSRIDGKILDRKSSGLDARATIAIAESSLTQGVATAP 122  
 ITA QLDGLF S GI YD R++G++L+ EK +GLD RAI+AI+AIESSLTQGVAA

-730-

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45

```

 Sbjct: 230 ITAEQLDGLNLSLGIHYDKERLNGKRLLSEWKLITGLDVRIVAIAIAMAESSLGTCQVAKSK 289
 Query: 123 GANMFGPAVDNNTINAGNPSDDKAVIKMTQRTTIQNQNTSPAIDQQAQPLSTGNIANA 182
 G+MFG+GA D N NA+ +SD+ A+ M ++TII N+N +F QD KA+ S G L+
 Sbjct: 290 GSNMFGYGAFDENPANAACKYSEDAIRHNVEDTIIANNKQTFERQDLAKAKWSLQQLDTL 349
 Query: 183 ARGQVYPTDASGSGKRRRAIMESIDKWDANGGISEISEKELLNTSSVAMGAVPTSYSVSR 242
 GQVYPTD SSG+RRA IM +D+WD HG +I +L TS VP Y S+
 Sbjct: 350 IDGQVYPTDTSGGGRRADIMTKLDQWIDDHGNTFDIPEHLKITSGTSPSEVFGYKRSQ 409
 Query: 243 ANQAGNYVAGTYPNQQRITW 261
 Y + TY +GQ TW
 Sbjct: 410 PQNVLITYKSETYPSQCTW 428

 A related DNA sequence was identified in S.pyogenes <SEQ ID 1987> which encodes the amino acid
 sequence <SEQ ID 1988>. Analysis of this protein sequence reveals the following:

 Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

 An alignment of the GAS and GBS proteins is shown below:

 Identities = 143/265 (53%), Positives = 184/265 (68%), Gaps = 5/265 (1%)

 Query: 2 VPSQPVVATTPQKSEVVTPA-----ITSGLDFOVAIPTAMASAAVYKHNGINDAYTHNL 56
 V + P + + Q R TP S +DL ++ IP+ AAYV+HW G +AYTH+L
 Sbjct: 135 VDTAPASSLSKQLPEARTPIQGLSPYVSDLDSEIDPSVNTYAAVYEHSSGKNAYTHKL 194
 Query: 57 LSRRYGIITAAQLDGLQSTGITVDSSRIDQKILDRKSSGLDARAIITAIARSSLGTC 116
 LS RYGI A Q+D +L+STGI YDS+RI+G+K+L EK SGLD RAI+AIR++SSLGTC
 Sbjct: 195 LSRRYGIKADQIDSYLKGSTGIAYDSTRNGKLLQWEEKSGLDVRAIVAIAIAMSSESLGTC 254
 Query: 117 GVATAPGANMFGGAVDNNNTINAGNPSDDKAVIKMTQRTTIQNQNTSPAIDQQAQPLST 176
 G+AT GANMFG+ A D + T A F+DD A++KMTQ+TII+N+N++FA-QD KA S
 Sbjct: 255 GIATLLGANMFGVAAFDDLTQASKFNDSIAIVKMTQDTIIRKNGSNFALQDLAKAFSR 314
 Query: 177 GNINVAARGGVYPTDASGSGKRRRAIMESIDKWDANGGISEISEKELLNTSSVAMGAVPT 236
 G LN A+ GGVYPTD +GSGKRRRA IME +DKWID HGG I EL SS + +VP
 Sbjct: 315 GQLNFASDGGVYPTDTTSGSKRRQIMEDLKNIDDHGGTFALPAELKQVSGSASFASVPA 374
 Query: 237 SYSVSPANQAGNYVAGTYPNQQRITW 261
 Y +S++ Y A +Y GQ TW
 Sbjct: 375 GYKLSKSYDVLGYQASSYAWGQCTW 399

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 50 Example 642

A DNA sequence (GBSx0682) was identified in *S.agalactiae* <SEQ ID 1989> which encodes the amino acid sequence <SEQ ID 1990>. Analysis of this protein sequence reveals the following:

55  
 60

```

 Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```



-731-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8627> and protein <SEQ ID 8628> were also identified. Analysis of this protein sequence reveals the following:

```

5 Lipop: Possible site: -1 Crend: 4
 McG: Discrim Score: 11.56
 GvH: Signal Score (-7.5): 0.870001
 Possible site: 27
10 >>> Seems to have a cleavable N-term signal seq.
 ALGM program count: 0 value: 11.88 threshold: 0.0
 PERIPHERAL Likelihood = 11.88 63
 modified ALGM score: -2.88

15 *** Reasoning Step: 3

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 8628 (GBS159) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 4; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 2; MW 41kDa).

GBS159-GST was purified as shown in Figure 198, lane 9.

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 643

A DNA sequence (GBSx0683) was identified in *S.agalactiae* <SEQ ID 1991> which encodes the amino acid sequence <SEQ ID 1992>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2668(Affirmative) < succ>
35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40 >GP:BA04699 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 32/76 (42%), Positives = 54/76 (70%)

 Query: 7 LGSVIELKNDSSQKVMITSRFFLYDNEGQLGYFDYSQCIFPISIVGNETYFENLEIDKVL 66
 +GS++ LK + K+MI +R P+ + G+ FDYSGC +P +V ++ ++FN E+ID+V+
 Sbjct: 4 IGSIVLKEGTSKLMIINRGPILEANGENRMFDYSQCFYPQGLVEDKVPYFNHSEINIDEV 63

45 Query: 67 FEGYYDENESSEMOKIP 82
 FEG+ D+ E+ QK+P
 Sbjct: 64 FEGFQDDEQRQPKLP 79

```

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 644**

A DNA sequence (GBSx0684) was identified in *S.agalactiae* <SEQ ID 1993> which encodes the amino acid sequence <SEQ ID 1994>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-14.81 Transmembrane 75 - 91 (69 - 99)
 INTEGRAL Likelihood =-14.38 Transmembrane 134 - 150 (129 - 179)
 INTEGRAL Likelihood = -8.49 Transmembrane 157 - 173 (151 - 179)
10 INTEGRAL Likelihood = -1.17 Transmembrane 50 - 66 (46 - 67)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.6922 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 645**

A DNA sequence (GBSx0685) was identified in *S.agalactiae* <SEQ ID 1995> which encodes the amino acid sequence <SEQ ID 1996>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.11 Transmembrane 40 - 56 (40 - 56)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1996 (GBS204) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 13; MW 32kDa) and Figure 53 (lane 2; MW 14.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 6; MW 39.7kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 646**

A DNA sequence (GBSx0686) was identified in *S.agalactiae* <SEQ ID 1997> which encodes the amino acid sequence <SEQ ID 1998>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 38
 >>> Seems to have no N-terminal signal sequence (or signal = aa 1-26)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-733-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:CA16670 GB:AJ302698 hypothetical protein [Staphylococcus  
haemolyticus]  
Identities = 60/254 (23%), Positives = 109/254 (42%), Gaps = 14/254 (5%)

Query: 2 VKSVSVSVGTQASTVAISMPSKVSALADAITKLSSFAEATLQGTAYSNKSYATGTLTP 61  
++ + + + Q+S V ++ S S + + F R+ LGS AY + K + + + F  
10 Sbjct: 3 IDMTVGKSKSQSDVGSTVKSISGGYDLSLQKQIMQFVGASELCQAYDSGKQFFSAVIA 62

Query: 62 MLQGMILFSETLSSEKTELTQTLVSVICGDEDLGSVVLSEKSLASDRASLKIAEALLHLND 121  
+ + + + E + C + + Y S + L L + + + EA+ L  
15 Sbjct: 63 LTBSIKTLQELTQACNDPVDQYQSEVDSQSLKESELLEDIEBLNKQISQLEMNASLKH 122

Query: 122 DFEPSKSAISSTGKNIKKLKKRIKSNQKKLDNMEFNHSAITVFADISNAQSTVWQALAA 181  
+ S + S I L + + + K + + KL L + F+A S + F + + + Q TV Q + +  
20 Sbjct: 123 KSSINSSLLSGNHQMISSLEQQKELEBKRLKQFDKASPNIFKEVESFQRTVQQINQ 182

Query: 182 VSTGPGSYNSKTGAPGPKPTSGGMWTKTVKMKEREDAVBELSKSKAKESKAKSKEN 241  
+ + + + F P MEW K + + E K + + + + K + + + K K SK +  
25 Sbjct: 183 AKT---AMDPGKQTFNIPAGKMEWAKVSCQKALE---VIMDKI-NQKAKDGKHLGNDI 235

Query: 242 TT-----KKSNV 248  
T KKSNI+  
Sbjct: 236 FTIIAYQQQKKSNV 249

No corresponding DNA sequence was identified in *S.pyogenes*.

30 SEQ ID 1998 (GBS270) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 2; MW 34.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 7; MW 59.2kDa).

The GBS270-GST fusion product was purified (Figure 206, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 265), which confirmed that the protein is immunoreactive on GBS bacteria.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 647**

40 A DNA sequence (GBSx0687) was identified in *S.agalactiae* <SEQ ID 1999> which encodes the amino acid sequence <SEQ ID 2000>. This protein is predicted to be outer surface protein F. Analysis of this protein sequence reveals the following:

Possible site: 23  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3323 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

-734-

SEQ ID 2000 (GBS316) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 3; MW 23kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 2; MW 41.8kDa).

GBS316-GST was purified as shown in Figure 206, lane 4.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 648

- A DNA sequence (GBSx0688) was identified in *S. agalactiae* <SEQ ID 2001> which encodes the amino acid sequence <SEQ ID 2002>. This protein is predicted to be actin-like protein arp3 (act4). Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0217(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 649

- A DNA sequence (GBSx0689) was identified in *S. agalactiae* <SEQ ID 2003> which encodes the amino acid sequence <SEQ ID 2004>. This protein is predicted to be diarrheal toxin. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.65 Transmembrane 65 - 81 (61 - 84)
INTEGRAL Likelihood = -3.98 Transmembrane 89 - 105 (85 - 106)

----- Final Results -----
bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB15175 GB:Z99120 alternate gene name: yueA-similar to
hypothetical proteins [Bacillus subtilis]
Identities = 452/1058 (42%), Positives = 664/1058 (62%), Gaps = 39/1058 (3%)

Query: 98 VMIFSITGYFMRKQYKQDLQERIDSHYDYLSDKSIIRLQKLAHQKRGQINHYPTISGL 157
+T+I S YF+++ Q K+ ++R Y YL +K RLQ LA++QK+ +H+P+ E +
Sbjct: 1 MTLTSTVQYFRDQGRKKRERKRVYKLYLQNKKKLQALAEKQKQVLEFHPFSTQDM 60

Query: 158 QEMADPTTHRIYKPTLPDPLFYRLGLGAVPTSYNTHYSQPERBQKK-DELENGYHLY 216
+ + RI+EK+ D+L RLQ G VP+SY I+ S + + + D L + + +
Sbjct: 61 KYLFSBISDRIWESLESKDYLRQLRIGTVPSPSYETINSGGLDANRDI DDLAKSQIMQ 120

Query: 217 FNNRYIKNMPIVANLGHGPGVYIGPGLVLEQLQLMVNCAPFPHSYHDVQFITIVPEEM 276

```

-735-

|    |        |                                                                  |      |
|----|--------|------------------------------------------------------------------|------|
|    |        | + I-N P+ +L+ GP+G +G +V + + + QI+FF-SYHD++F+ I ESE               |      |
|    | Sbjct: | 121 RVYKDIRNAPVTVDLABGFMGLVGKSQIVNWEIHQILGQLSPFNSYHDLRFVPTFHSEBY | 180  |
| 5  | Query: | 277 DWSNWRWLPHETLQDVRVGRGVYVNRSDQVLSNLIQLKLRTRQRBHKSAGBTLPSP     | 336  |
|    |        | W NM+ +P + + +GF+YN+++RDQ+L+SL +++++ +R+ + KR F P P              |      |
|    | Sbjct: | 181 KDWSNWKCVQPFQMFIYAKGFITYNEQTRQLLSSLYELIR---ERDLEDDKRLQFKP    | 236  |
|    | Query: | 337 HYVVITDEKLILDRVIMSEPTTDFTELGCSLIFVQDVMSSLSHNIKITIINKRNTGQL   | 396  |
|    |        | H+V +T+++LI +HVI+E+ LG S I + SLSENI T+ + + G +                   |      |
| 10 | Sbjct: | 237 HFVFTVITNQLLSERHVLVLEBQSHLIGLSTIVAARTKESLSENITLIVRYINBISDIT  | 296  |
|    | Query: | 397 VIEBGLKATDFELDHFLDYDKNENSRRLAPAHNLQNLKSIPSAVTVMHVAQEFED      | 456  |
|    |        | +I+ + F LDH + D E SR L LNH + +SIP V+F+E++ A+E+ +                 |      |
| 15 | Sbjct: | 297 LIQKKAVRIPFRLDHQRE-DNERFSRTLTTLNHQVGITNSIPETVSPLELFAKEVKE    | 355  |
|    | Query: | 457 LHVQERNISHPYKSSAVPLGLQDQDVIYMLHKAHQPGLVAGTOSGKSELIQSYI       | 516  |
|    |        | + +Q+RN++ KS +VP+G +G+DVIYMLHKAHQPGL+AGTOSGSE +Q+YI              |      |
|    | Sbjct: | 356 IGIQCRNLTSBSSKSLSPVIGYKGDVIYMLHKAHQPGLLAGTOSGSEFLQTYI        | 415  |
| 20 | Query: | 517 LSLAVNPHDVAFLIIDYKGGNANLFDLPFLHLLGTTITNLGQA--SMRALVSTINSLK   | 574  |
|    |        | LSLAV+PHF+ AFLIIDYKGGNA F+++PHLLGTTIN++G+ SMRAL SI +ELK          |      |
|    | Sbjct: | 416 LSLAVNPHFHEAFLIIDYKGGNAQPPFKTIPHLGTTITNIESSKNFMRALASIKSEL    | 475  |
|    | Query: | 575 RRGLFAKADVNHNQYQKCYKLGVESEMPHPLFLISDEFARLKSQPEPFKEINVESTARI  | 634  |
|    |        | +RQRL + VNHN Y K YK G+ MPHEFLISDEFARLKS +P+P+ELNS ARI            |      |
| 25 | Sbjct: | 476 KRQLFDQYQVNHINDITLYKQKAEVAMPHPLFLISDEFARLKSSEFPRELVAASARI    | 535  |
|    | Query: | 635 GRSLLGHLILATQKPSGVDDQIWSNSRFLALKAVRAGDSNEMHTPDAEITQAGRAY     | 694  |
|    |        | GRSLG+HLILATQKP G++DDQIWSNSRFL+ALKV D DS E+L DAA IT GR Y         |      |
| 30 | Sbjct: | 536 GRSLGHLILATQKPGGIIIDQIWSNSRFLKALKVQDATDSKELKNSDAANTVTGRGY    | 595  |
|    | Query: | 695 LQVGNNEVVELFQSANSGADYQPEKDDQGIEDHTIYSINDLGQYRIINDLDSQLDAQNI  | 754  |
|    |        | LQVGNNEVVELFQSANSGA Y E G E D I + D G LS +D +N                   |      |
| 35 | Sbjct: | 596 LQVGNNEVVELFQSANSGAPYLEEV--YGTDE-IAIVITDTGLI-----PLSEVDTONA  | 647  |
|    | Query: | 755 -KEVPTFLDAIVENIQALTKEMGISLQPMFLPPLSNQIAVTDLRKSGSVLMWSKAPYK   | 813  |
|    |        | K+V TE++A+V+ I+ + EMGI LP PWLEPL+ +I T L+                        |      |
|    | Sbjct: | 648 KKVQCTEIRAVVDEIERIQDEMGIKLPSTWLPPLAERIPT-----LPSINEKDH       | 698  |
| 40 | Query: | 814 AVLGFMDIPSCQPEVAYHDFEDGHLISIFAGPSMGKSTALQTVINDLARNSEPFILNLY  | 873  |
|    |        | ++D P Q Q + +DG++ IF GKS A T M A + PE L+Y                        |      |
|    | Sbjct: | 699 PHFAYVDEPDLQRAPIAYKMMEDXNIGIFGSSGYKSLAAATFIMSFAIVTYPEELHVV   | 758  |
|    | Query: | 874 LFDGFGNLLPLRLRLPHVADFFITDDKEIAKFIARIKVMMSDRKKALSRYNVATAKLYR  | 933  |
|    |        | +FDG L LPL +LPH AD+F D KI KF+ RIK E+ R+ + K+Y                    |      |
| 45 | Sbjct: | 759 IFDFGNGTLLPLAKLPETADYFMDQSRKIEKFMIRIKKEIDRRKLPEKEKLSHLYN     | 818  |
|    | Query: | 934 QVSGETMPQILIVISYEGLEEAQFTINLEACPQNISRQDSSLGISLVISAGRTAALRSS  | 993  |
|    |        | +S E +P I I ID++ + + LE+ F +SRDG SLGI + +A R A+R S               |      |
| 50 | Sbjct: | 819 ALSERELPFIFTITDNDVIDEN--HELESFVQLSRDQSLGTYMLTATRVNVRG        | 876  |
|    | Query: | 994 IMANLKERIALKLTDDSESPTLVGRHQHIMEDIPEGGLIKRDDIEVLQVALSTGRTFTD  | 1053 |
|    |        | L+ NLK +I L D SE ++ GR + +E IPGR +I++++ Q+ L + +                 |      |
|    | Sbjct: | 877 LLNLEKTKIVHYLMDQSEGSYIGRPFKNLEPIPGRVITQKEELVFAQMFLPVDADDDIG  | 936  |
| 55 | Query: | 1054 IINNQNESDAMNSKWTG-PRPKAIPVPEELTFDDPMATDSVQADLSANRL--PLGLEM  | 1110 |
|    |        | + N + + + + + + + + P IP+PE L+ + S++ L L P+GL                    |      |
|    | Sbjct: | 937 MPNELKSDVQKLGFRFASMEQAPITMPLPESTLSTRL---SIRFKLERKELSPVIGLER  | 992  |
| 60 | Query: | 1111 VDVSYSLALNRFKRMLYMSDSDESLEAVGSHIKVLL 1148                   |      |
|    |        | V L + KH L + + +++Kv+L                                           |      |
|    | Sbjct: | 993 ETVSVPVFDLGHKHKCLILGQTRG---KTNVLKVM 1026                     |      |

There is also homology to SEQ ID 24.

- 65 A related GBS gene <SEQ ID 8629> and protein <SEQ ID 8630> were also identified. Analysis of this protein sequence reveals the following:

-736-

Homology to a bacterial toxin

The protein has homology with the following sequences in the databases:

&gt;OMNI|NF01BS3725 diarrheal toxin

Score = 203 bits (511), Expect = 4e-51  
 Identities = 123/377 (32%), Positives = 198/377 (51%), Gaps = 22/377 (5%)

Query: 1 MGSLDLFQFWLPLPLNQIAVTDLRKEGSDVLMNSKAPSYKAVLGSMDIPSQQAQEVAYHDF 60  
 MGI LP PHLPLP +I T L+ +D P Q Q +  
 Sbjct: 704 MGIRKLSPFWLPLASRI PRT-----LPSNKKDHFAFVDFEPLQRQAPIAYKM 754

Query: 61 RDDGHLISIFAGPSMGKSTALQVTMDLARHNSPFINLYLFDFGTGGLPLRLRHVADF 120  
 +DG++ IF GKS A T M A +PE L++Y+FDG LLPL +LPH AD+  
 Sbjct: 755 MEDGNIGIFGSSGYGKSLAAATFLMSFAIVYTPPEELHVTYDFGNGTLLPLAKLPHADY 814

Query: 121 FTIDDEKIAKFIARIKVRMSDRKALSRYNVATAKLYRQVSGSTMPQLILVIDSYEGLR 180  
 F +D KI KP+ RIK E+ RK+ ++ K+Y +S E +P I I ID+++ ++  
 Sbjct: 815 FIMDQSRKIEKFMIRIKESIDRRKRLFRKKIISHIKMYNALSEBELPFPIITIDNFDIVK 874

Query: 181 EAQTPTNLEACPQNI SRDQSLGISLVISAGRTAARLSSIMANLKERIALKLTUDESRT 240  
 + LE+ F +SRDG SLGI +++A R A+R SL+ NLK +I L D SE +  
 Sbjct: 875 DEM--HELESEFVQLSRDQSLGIYFMLTATRVNAVRQSLNNLKTIVHYLMDCQSEGY 932

Query: 241 LVGRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFDIINNIQNESDAMNSKWTG-PR 299  
 + GR + +E IPR +I++++ Q+ L + + + N +++ + ++ +  
 Sbjct: 933 IYGRPKFNLEPTPRVIIQKEELYFAQMF LVDADDIGMFNELKSDVQKLGKRPASMEQ 992

Query: 300 PKAIPVPEELTFDDPMATDSVQADLSANRL--PLGLENVDSYSLALNRFKHMLYMSD 357  
 P IP++PE L+ + S++ L L E+GL V L + KH L +  
 Sbjct: 993 PAPFPHLPESLSTREL----SIRFKLERKPLSVPTGLHEETVSPVYFDLGHKHLCLIQ 1048

Query: 358 SDESLEAVGSHIIRVLL 374  
 + +++K+V+L  
 Sbjct: 1049 TGRG---KTNVLKVL 1061

SEQ ID 8630 (GBS326) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 5; MW 66kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 5; MW 91kDa).

GBS326-GST was purified as shown in Figure 212, lane 5.

GBS326LN was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 2-4; MW 114kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 6; MW 114kDa). The purified protein is shown in Figure 236, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 650

A DNA sequence (GBSx0690) was identified in *S. agalactiae* <SEQ ID 2005> which encodes the amino acid sequence <SEQ ID 2006>. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2693 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 651

A DNA sequence (GBSx0691) was identified in *S.agalactiae* <SEQ ID 2007> which encodes the amino acid sequence <SEQ ID 2008>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3933 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 652

A DNA sequence (GBSx0692) was identified in *S.agalactiae* <SEQ ID 2009> which encodes the amino acid sequence <SEQ ID 2010>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -7.32 Transmembrane 225 - 241 (219 - 246)

----- Final Results -----
bacterial membrane --- Certainty=0.3930 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BA04693 GB:AP001510 unknown conserved protein [Bacillus halodurane]
Identities = 83/320 (25%), Positives = 162/320 (49%), Gaps = 1/320 (0%)

Query: 103 VNFILNPSNLFITNATAKIAVRSLEGIMRPEKPGPSEFLYQKCFVFALLTQHDVRLY 162
 ++ I+ P N+ ++ + + P + PE ++ + LL + . Y
Sbjct: 106 LHLIVSPENVLSXGLDVTIFHYGVKDSIPPYETDPERLFLEIRATILWLIDGHRHFREY 165

Query: 163 NGALSVIEVSDFLKSIYHAKTIAQVRLDITIDYBQQVEVETHLAKVSRAKYKLYKYSV 222
 ++S KS+ T++ +R++I + Q+ E + L KV + K+ + K+
Sbjct: 166 MNYHTLKLSPKAKSLVQQTTLBGLKELR-HWIQEHEQKQLHKVPKTKNTIQKNAGI 224

Query: 223 WEGALSTILLPLVYLWFIHNPFKKMLAATSPFKVDYNQVINKLHVVKVSKLEPYTKY 282
 L A +I +VY++ P +E A+ ++ +Y+QVI+ LE + +P KY
Sbjct: 225 GLIAALVPAITTYIVVLAFLQPRGKAPTASHAAYLNKNTYSQVIDLPLSPNSMPPVVKY 284

Query: 283 ELAYSINGMSEFSEQREVLNNVILTKIDELYLDYWINIGRGLDDDAIDAARLDDSDLV 342
 +LA SY+ RE + N + L+ E Y DYWI IGRG ++ AID A+ L D +
Sbjct: 285 QLAQSYVAIEPLQAYERTLNKNVILQAARSYPDYWIAIGRGRNEKAIDIAIQLQKQEW 344
```

-738-

Query: 343 IYAIVQKMDQVRKDNLSGKDKREKQLSELQUTUYDKYKWKRTALTDERSKSKNSNNHSTM 402  
 +YA V++ ++V+ D +LGGK+RE + E++ + D Y ++ + + E+ N+ ++M  
 Sbjct: 345 YIANVKKREEVKSDDNLSGKEREDLTKREAEIIDDYMRLEELAEGBGAQPNAPAEAM 404

Query: 403 SNKSSSSSSSTTASTSSKTK 422  
 +E + S + + K  
 Sbjct: 405 ELAEDEGUTEEDDSNQERAK 424

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 SEQ ID 2010 (GBS337) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 3; MW 50.3kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 653

- 15 A DNA sequence (GBSx0693) was identified in *S.agalactiae* <SEQ ID 2011> which encodes the amino acid sequence <SEQ ID 2012>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have a cleavable N-term signal seq.  
 INTSEGAL Likelihood = -14.01 Transmembrane 131 - 147 ( 122 - 153)

20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6604 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 8631> which encodes amino acid sequence <SEQ ID 8632> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8  
 MoG: Discrim Score: 13.38  
 GvH: Signal Score (-7.5): -1.25

30 Possible site: 23  
 >>> Seems to have a cleavable N-term signal seq.  
 ALOM program count: 1 value: -14.01 threshold: 0.0  
 INTSEGAL Likelihood = -14.01 Transmembrane 127 - 143 ( 118 - 149)

35 PERIPHERAL Likelihood = 16.13 113  
 modified ALOM score: 3.30

\*\*\* Reasoning Step: 3

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6604 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8632 (GBS140) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 3; MW 43kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 8; MW 18kDa).

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.



**Example 654**

A DNA sequence (GBSx0694) was identified in *S.agalactiae* <SEQ ID 2013> which encodes the amino acid sequence <SEQ ID 2014>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1486(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 655**

A DNA sequence (GBSx0695) was identified in *S.agalactiae* <SEQ ID 2015> which encodes the amino acid sequence <SEQ ID 2016>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-14.59 Transmembrane 984 -1000 (976 -1009)
 INTEGRAL Likelihood = -9.71 Transmembrane 19 - 35 (15 - 42)
 INTEGRAL Likelihood = -9.50 Transmembrane 872 - 888 (865 - 890)
 INTEGRAL Likelihood = -6.37 Transmembrane 927 - 943 (924 - 951)
 INTEGRAL Likelihood = -4.19 Transmembrane 831 - 847 (828 - 847)
 INTEGRAL Likelihood = -2.87 Transmembrane 899 - 915 (899 - 916)

----- Final Results -----
bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8633> which encodes amino acid sequence <SEQ ID 8634> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1 Crend: 6
SRCFLG: 0
McG: Length of UR: 20
Peak Value of UR: 3.40
Net Charge of CR: 3
McG: Discrim Score: 13.67
GvH: Signal Score (-7.5): -3.27
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOW program count: 6 value: -14.59 threshold: 0.0
 INTEGRAL Likelihood =-14.59 Transmembrane 973 - 989 (965 - 998)
 INTEGRAL Likelihood = -9.71 Transmembrane 8 - 24 (4 - 31)
 INTEGRAL Likelihood = -9.50 Transmembrane 861 - 877 (854 - 879)
 INTEGRAL Likelihood = -6.37 Transmembrane 916 - 932 (913 - 940)
 INTEGRAL Likelihood = -4.19 Transmembrane 820 - 836 (817 - 836)
 INTEGRAL Likelihood = -2.87 Transmembrane 888 - 904 (888 - 905)
PERIPHERAL Likelihood = 3.82 936
modified ALOW score: 3.42
icml HYPID: 7 CFP: 0.684

*** Reasoning Step: 3

```

-740-

----- Final Results -----

bacterial membrane --- Certainty=0.6838 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AB86324 GB:AE000938 phage infection protein homolog

[Methanothermobacter thermoautotrophicus]

Identities = 96/454 (21%), Positives = 190/454 (41%), Gaps = 63/454 (13%)

Query: 1 MLKIKYILGRIMKR-NNFRILWYIIAVALFVAIAGINLKLQSDHAKENKTTQSATNTKL 59  
 M K I + M K N ++ ++IAV + + A+ + +Q ++T+ +  
 Sbjct: 1 MRKALEIFWKDMKTVKNSPVVLFVIAVIIICIPALYAV-FNIQATLDPSYRTSS-----I 53

Query: 60 NIALVNEQVNSNGKRSYNLGSYIKSIERINSCNWSVVSRTAQNGLDKGYQLMVIIP 119  
 +A+VNEQ N+GA ++ + ++ +W V R A +GL KG Y ++IIP  
 Sbjct: 54 EVAVVNEQMGADFNGTHLVNGAEFVSELKRNRFWDQFVDRSDAMDGIRKGYAVLTIIP 113

Query: 120 NNFSQKLLDVNKANABQTTISYKVNAGNLALEKKATEKEKDIVSELNSHLVNNYMASIL 179  
 NFS LL + Q +I Y VN K N + + ++NS +V +  
 Sbjct: 114 GNFSOLLISIKNGTFRQASIKYVNDKLNPAVRITNAGADALQAKINSEVVKITDGIIV 173

Query: 180 SNLYTAQENVQA-----MVNVQSGNISNYQNLDSATNF--QNIFPAL----- 221  
 + A E +A VN +GN+ + L + ++ QN++ +L  
 Sbjct: 174 GKISEAGELANRNDILRTKRFVNLKGNLQKIDETLSTANSLEKGNLASSLKTULP 233

Query: 222 -VNQSSSSITANESLKKS-----LEASDNMFNDLVTTQTNTGKDLSSL----- 263  
 + +++ + SL +S +++ ++ ++ +I+ L+SL  
 Sbjct: 234 EIRDANFVKEKYSLSLEYIGKDPKALSTVQSMESHLSEATSMKYLRAVLASLSATG 293

Query: 264 -----IEQRHQDSISYEAFSTSLLENNELLEKQLSDIITQAKDQETLSSQLNSNG 316  
 I+Q + + L + ++L K +D I + + + S LM +M  
 Sbjct: 294 DPKLKTADIDQIDTNIKASSVLGILQTIESDLKNGTTRIVKLKASIDRMSALANKMD 353

Query: 317 D-DNNHNHKNSSAYLNVARQKIQELSEALKSQDNIADQSQEQLDKIVREGLASYFAKRN 375  
 D +++SA L +A + + A+ +D S +L+ I + L S + +  
 Sbjct: 354 SRDEIDAMQDASAKLGIANARWFTWRSAL-----QDASRKLMTSDDDLNLSLVKLAD 406

Query: 376 KDNITLLELLKSHSTNEK----TLKDPKAKVADP 405  
 D + E +S EK +K++ + +A F  
 Sbjct: 407 IDPSAVREYFRSPVMEKEHIYPVKYNGSLAPF 440

SEQ ID 8634 (GBS250) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 4; MW 136kDa).

GBS250-GST was purified as shown in Figure 203, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 66

A DNA sequence (GBSx0696) was identified in *S.galactiae* <SEQ ID 2019> which encodes the amino acid sequence <SEQ ID 2020>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5009 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-741-

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA6375 GB:X65276 ORF1 [Clostridium acetobutylicum]  
Identities = 35/91 (38%), Positives = 53/91 (57%)

5 Query: 1 MAQILTPPEELRSSAQKTYAGSCQVTEVLNLLTQSDAVIDENNDGSTFDSFPAQFNEELSP 60  
MAQI +TPPEEL+S AQ Y +++ + + + + I E W G F ++ Q+N+L  
Sbjct: 1 MAQISVTPPEELKSGAQVYTIQSKEIDQATQKVNMSNSTIABEWKGAQFQAYLEQYNQLHQ 60

10 Query: 61 KITEPAQLLEDINQQLKVDIIBQTDADIA 91  
+ +F LLS +NQQL K AD + + DA A  
Sbjct: 61 TVVQFENLLESVIMQQLKKYADTVAERDAQDA 91

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
15 vaccines or diagnostics.

#### Example 657

A DNA sequence (GBSx0697) was identified in *S.agalactiae* <SEQ ID 2021> which encodes the amino acid sequence <SEQ ID 2022>. Analysis of this protein sequence reveals the following:

20 Possible site: 22  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3741 (Affirmative) < succ>  
25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
30 vaccines or diagnostics.

#### Example 658

A repeated DNA sequence (GBSx0698) was identified in *S.agalactiae* <SEQ ID 2023> which encodes the amino acid sequence <SEQ ID 2024>. This protein is predicted to be carbamoylphosphate synthetase (carB). Analysis of this protein sequence reveals the following:

35 Possible site: 23  
>>> Seems to have a cleavable N-term signal seq.  
INTEGRAL Likelihood = -1.33 Transmembrane 807 - 823 ( 807 - 823)

----- Final Results -----  
40 bacterial membrane --- Certainty=0.1532 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

45 >GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus  
lactis]  
Identities = 771/1062 (72%), Positives = 901/1062 (84%), Gaps = 5/1062 (0%)

50 Query: 1 MPERTDIRKINIVIGSGPTVIGQAARFDYSGTQACLSLKEBGGYQVVLVNSNPATIMTDKDI 60  
MPER DI+KIM+IGSGPT+IGQAARFDY+GT+ACL+LKEBGGY+VVLVNSNPATIMTD++I  
Sbjct: 1 MPKRNIDIKIMITIGSGPLIIGQAARFDYAGTACIALKLEBGGYVVLVNSNPATIMTDREI 60

|             |                                                                                                |                                                                             |      |
|-------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|------|
| Query: 61   | ADKVITEPTLEFVTRILRKERFDALLTLOGQTOLANMAMALEKNGILEELAVELLOTKL                                    | 120                                                                         |      |
| Sbjct: 61   | ADVYITEPTLEFV++ILRKERFDALLTLOGQTOLANMAM L&K GILEELAVELLOTKL                                    | 120                                                                         |      |
| 5           | Query: 121                                                                                     | SAIDKAERDRLFKQLMEELANQPIPESSIVNSVEAIFQARQIGYPLIVRAPPTLGTQGG                 | 180  |
|             | SAID+AEER+L&K+L E +N+P+ S+I +VEEA A++ICYP+IV PAPT+QGNQGG                                       |                                                                             |      |
| Sbjct: 121  | SAIDQAERDLFKELCESINEPLCASDIATTVEBAINIADKIGYPIVGPAPTGMTQGG                                      | 180                                                                         |      |
| 10          | Query: 181                                                                                     | MCNQEQIANDITTKGLKLSFVTCCLIRRSIAGFKETIYSVMRDAANNAIVVCMENFDPV                 | 240  |
|             | +CD +E+L +I GLKLSFVTCCLIE SIAG+KEIYYSVMRD+ADNA+VVCMMENFDPV                                     |                                                                             |      |
| Sbjct: 181  | ICUTEELREIVANGKLSFVTCCLIRRSIAGFKETIYSVMRDSANNAIVVCMENFDPV                                      | 240                                                                         |      |
| 15          | Query: 241                                                                                     | GHTGDSIVFAPACTLSDVENQILRDASLDIIRALKIBGGCNVQLALDENSFKYYVIEVN                 | 300  |
|             | G+HTGDSIVFAP+QTLSD E Q+LRDASL+IIRALKIBGGCNVQLALDENS++Y VIEVN                                   |                                                                             |      |
| Sbjct: 241  | GVTGDSIVFAPSGTLDSEYQMLRDASLNIIRALKIBGGCNVQLALDENSFKYYVIEVN                                     | 300                                                                         |      |
| 20          | Query: 301                                                                                     | PRVSRSSALASKATGYPIAKLAAKIAGLTLDEVINPITKTTYAMPFALDYVAKMRFP                   | 360  |
|             | PRVSRSSALASKATGYPIAK++AKIA+G+TLDE+INP+T TYAMPFALDYVAK+ RF                                      |                                                                             |      |
| Sbjct: 301  | PRVSRSSALASKATGYPIAKMSAKIAIGMTLDEINPVTNKTYAMPFALDYVAKIARF                                      | 360                                                                         |      |
| 25          | Query: 361                                                                                     | PFDKFBGDRKLGTQMKATGEVMAIGRNIESLLKAVRSLEIGVPHI KPADLNNSDVL                   | 420  |
|             | PFDKFB+CDR LGTQMKATGEVMAIGRNIESLLKA RSLEIGV H++ +D+ L                                          |                                                                             |      |
| Sbjct: 361  | PFDKFBGDRHLGTQMKATGEVMAIGRNIESLLKAVRSLEIGVPHNEMTATRADDEKL                                      | 420                                                                         |      |
| 30          | Query: 421                                                                                     | LEKIRKADDRLPYIARAIPEHYSTEKIASITSDFPDLKRVIVLELDLISKNRLDIN                    | 480  |
|             | KK+ K +DDRFPY++EA+RE IE+A+L LT ID PFLDKL IVE+E+ L N +                                          |                                                                             |      |
| Sbjct: 421  | YEMKVTQDDRFPYVSEAIRRGPIEBIAULTKIDPFLDKLYIVTEINQLKVNTEPEE                                       | 480                                                                         |      |
| 35          | Query: 481                                                                                     | ILKKVINGKGFSDKAIASLQINQEDVRMRKEAGLIPVYIMVDTCASEFSDATPYFSTY                  | 540  |
|             | +LK K GFSD+ IA LN + ++VR R+E I+PVYIMVDTCA+EF+S+TPYFSTY                                         |                                                                             |      |
| Sbjct: 481  | LKATKNGFSDREIAKLMNVTPERVRRRQENKIIPVYIMVDTCAEFESSTPYFSTY                                        | 540                                                                         |      |
| 40          | Query: 541                                                                                     | AVENESLISDKASILVLGSGPIRIGCGVEFDYATVHSVKAIRSGEPAIVINSNPETVST                 | 600  |
|             | ENES SDK I+VLGSGPIRIGCGVEFDYATVH VKAI+ G EAI++NSNPETVST                                        |                                                                             |      |
| Sbjct: 541  | EWENESKRSDKEKILVLGSGPIRIGCGVEFDYATVHCVKAIQALGKEAIVINSNPETVST                                   | 600                                                                         |      |
| 45          | Query: 601                                                                                     | DFSISDKLYFEPLTFEDVMNVLDL&KPSGVILQFGQQTAINLAKDLNKAQVKILGTQLED                | 660  |
|             | DFSISDKLYFEPLTFEDVMNVLDL&P VI+QFGQQTAINLA+ L+KAGVKILGTQ+ED                                     |                                                                             |      |
| Sbjct: 601  | DFSISDKLYFEPLTFEDVMNVLDLEPLVTVIQQGQTAINLAELH&KAGVKILGTQVED                                     | 660                                                                         |      |
| 50          | Query: 661                                                                                     | LDR&ENRKQFEATLQALNIPQPPGATATBEEAVNAQKIGYPVLVRPSFVLGRANKIV                   | 720  |
|             | LDR&E+R FE LQ L+IQQPPG TAT BEEAV A KIGYPVL+RPS+VLGRAM+I+                                       |                                                                             |      |
| Sbjct: 661  | LDR&ERDL&FKALQDLIDIPQPPGATATBEEAVANANKIGYPVLIRPSFVLGRAM&EII                                    | 720                                                                         |      |
| 55          | Query: 721                                                                                     | ENEDRLHYMTAVKASPD&PVLIDAYLIGK&CEVD&ISDGNILIPGIMEHIER&GVHS                   | 780  |
|             | NE+DLR YM AVKASP+P&VL+D+YL G+CEVD&I DG++LPGIMEHIER+GVHS                                        |                                                                             |      |
| Sbjct: 721  | NNE&DLRDMNR&KASPD&PVLVD&YLQ&Q&CEVD&ICDG&VELLPGIMEHIER&GVHS                                     | 780                                                                         |      |
| 60          | Query: 781                                                                                     | GDS&AVYPPQTLSETIETIVDYTKRLAIGLNCIG&NNIQFVILQDKQVYVIEVNRASRT                 | 840  |
|             | GDS&AVYFPQ L&+ II+TIVDYTKRLAIGLNCIG&NNIQFVI +++VYVIEVNRASRT                                    |                                                                             |      |
| Sbjct: 781  | GDS&AVYPPQL&SIIDTIVDYTKRLAIGLNCIG&NNIQFVIEBQVYVIEVNRASRT                                       | 840                                                                         |      |
| 65          | Query: 841                                                                                     | LPFL&KVTHIPMAQVATKVL&GLCNFTGYDLP&SDMVH&KAPF&SPTKLAKVD&SL                    | 900  |
|             | +PFL&KVTH+IPMAQ+AT++ILG+ L + Y L P DMVH+KAPF&SPTKLAKVD&SL                                      |                                                                             |      |
| Sbjct: 841  | VPFL&KVTHIPMAQVATQML&GNL&GLGY&GLAFT&DMVHV&KAPF&SPTKLAKVD&SL                                    | 900                                                                         |      |
| 70          | Query: 901                                                                                     | CEPMK&TE&GV&SDINLQ&ALY&K&P&A&Y&L&H&P&DY&Q&NIV&F&TV&D&DK&E&L&E&L&K&Y&Q&SI    | 960  |
|             | CEPMK&TG M&SD+ L+K&LY&K&P&A&Y&L&H&M&DY&G+++FTV D DK&E L L&K + I                                |                                                                             |      |
| Sbjct: 901  | CEPMK&T&L&M&SDVIL&K&LY&K&P&A&Y&L&H&M&DY&G&V&L&F&V&A&D&E&K&E&L&L&A&D&F&E&I                      | 960                                                                         |      |
| 75          | Query: 961                                                                                     | CYRIYATQCTAIYF&ANG&LE&TVL&G&L--&G&EN&N&H&I&POLI&N&K&GI&Q&AVINT&Y&Q&GN&N&ID- | 1017 |
|             | G&Y + AT G&Y + N&G L V K&L GE++ + + I+ G++Q&AV+NT+G                                            |                                                                             |      |
| Sbjct: 961  | C&Y&L&V&AT&G&T&A&F&L&K&E&N&G&LY&V&R&E&K&L&G&G&E&B&E&G&L&V&E&D&I&R&Q&R&V&Q&AV&N&T&G&N&T&R&A&S&L | 1020                                                                        |      |
| 80          | Query: 1018                                                                                    | --NID&ALI&R&R&S&A&I&G&Q&V&F&L&T&S&L&T&A&H&M&F&V&L&E&S&A&P&T&L&K 1057        |      |
|             | D IR+ AI +G+PL&T&S&L&T A+ KV++SR+PT K                                                          |                                                                             |      |
| Sbjct: 1021 | T&T&A&I&D&G&F&R&I&R&Q&E&A&I&S&R&G&I&P&L&T&S&L&T&V&A&I&L&K&V&M&Q&R&S&P&T&R 1062                 |                                                                             |      |

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2025> which encodes the amino acid sequence <SEQ ID 2026>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence
5 INTEGRAL Likelihood = -1.17 Transmembrane 773 - 789 (773 - 789)

----- Final Results -----
bacterial membrane --- Certainty=0.1468 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
lactis]
15 Identities = 753/1030 (73%), Positives = 876/1030 (84%), Gaps = 6/1030 (0%)

Query: 1 LALKEBQYKVLVNSNFATIMTDKEIADKYVIEPTLEFVNRIIRKRPDAILPTLGGQT 60
LALKEBQY+V+LVNSNFATIMTD+EIAD VYIEP+TLEFV+I+RKRPDA+LPTLGGQT
35 Sbjct: 35 LALKEBQYEVVLVNSNFATINTEREIADTVYIEPTLEFVSKILKKRPDALLPTLGGQT 94

20 Query: 61 GLNVMALSKAGILDDLEIELLCTKLSAIDQAEADRDLFKQLMCBQDPIPESTIVKTVD 120
GLNVM LSK GIL++L +ELLCTKLSAIDQAE+LFK+L ++++P S I TV+E
Sbjct: 95 GLNVAEILSKTGILLEEINVELLCTKLSAIDQAEDELFKCLCSINPLCASDIATVVE 154

25 Query: 121 AVTFARDIGYFVIVRPAPTLGOTGGGICSSSEELCEITENGLKLSFVTOCLIEBSIAGFK 180
A+ A IGYP+IV PAPT+GGTGGGIC+EEEL EI NGLKLSFVTOCLIE SIAG+K
Sbjct: 155 AINADIKGYPIIVGPAPTMGTGGGICDTSEELREIVANGLKLSFVTOCLIEBSIAGXK 214

30 Query: 181 EIEYEVMRDSADNALVVCNMENFDVGIHTGDSIVFAFTQTLSDIEQMLRDLASKIIRA 240
EIEYEVMRDSADNA+VVCNMENFDVGI+HTGDSIVFAF+QTLSD E QMLRDLAS IIRA
Sbjct: 215 EIEYEVMRDSADNAIVVCNMENFDVGIHVITGDSIVFAESQTLSDNEQMLRDLASINIRA 274

35 Query: 241 LKIEGGCNVQLALDPYSFKYYVIEVNFVRSSSALASGATGYPIAKLAAKIAGVLTDEM 300
LKIEGGCNVQLALDP S++Y VIEVNFVRSSSALASGATGYPIAK++AKIA+G+TLDE+
Sbjct: 275 LKIEGGCNVQLALDPNSYSEYRVIEVNFVRSSSALASGATGYPIAKMSAKIAGVLTDEI 334

40 Query: 301 INPITGTTAMFEPALDYVVAKIPFPFDKPEHGRQLGTQNKATGEVMAIGRNLESLL 360
INP+T TYAMFEPALDYVVAKI RFPFDKFE+G+R LGTQNKATGEVMAIGRN+ESLL
Sbjct: 335 INPVTNKTYAMFEPALDYVVAKIARFPFDKPFNGDRHLGTQNKATGEVMAIGRNLESLL 394

45 Query: 361 KACRSLEIGVCHNEMTSLNISDEELVTKVIKQDRLFYLSAIRRGYSIELESLLTKI 420
KA RSLEIGV HNEMT DE+L K++K QDRLFY+SEAIRRG IBE+ LTKI
Sbjct: 395 KAVRSLEIGVPHNEMTEATRADOELKYRHMVTKQDRLFYLSAIRRGIPETRIADLTCKI 454

50 Query: 421 DLFFLDKLLHIVEIBQELQMIVDHLSELKAKRYGFSQDKIARIWQKSDIRAMRHSHS 480
D+FFLDKLL+IVEIE +L++++ E LK A+ GFSD++LA++W ++R R +
Sbjct: 455 DIFFLDKLLYIVEIENQLKVNIPFELLLKTAQKGFSDRIAKLVNVTPEEVRRRQENK 514

Query: 481 LYPVYKMDVTCARFPDAKTFYFYSTYELNBSVOSNKESILVLSGGPIRIGQGVEDPYAT 540
+ PVMYKMDVTCARFP++ TPFFYSTYE ENBS +S+KR I+VLSGGPIRIGQGVEDPYAT
Sbjct: 515 IIPVYKMDVTCARFPESSTPYFFYSTYENBSKRSDEKILVLSGGPIRIGQGVEDPYAT 574

55 Query: 541 VHSVKAIKAGYERAIMNSNPETVSTDFSVSKLYFELPLTFEDVMNVIDLEQPKGVIVQF 600
VH VKAIQ G EAI++NSNPETVSTDFSVSKLYFELPLTFEDVMNVIDLE+P VIVQF
Sbjct: 575 VHCVKAIQGLKGAIVINSNPETVSTDFSVSKLYFELPLTFEDVMNVIDLEELPLVIVQF 634

Query: 601 GGQTAINLAQALSEAGVTILGTQVEDLDRADRDLFEKALKEIGIPOGQGTATNEERL 660
GGQTAINLA+ LS+AGV ILGTQVEDLDRADRDLFEKAL++L IPOQ G TATNEER+
Sbjct: 635 GGQTAINLAHLSKAGVKILGTQVEDLDRADRDLFEKALQDLDIPOPGATATNEERAV 694

60 Query: 661 EAAKKIGYFVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPVLVDSPYIGKE 720
A KIG+PVL+RPS+VLGGRAMEI+N+DLR+Y+ AVKASPEHP+LVDSY+ G+KE
Sbjct: 695 ANANKIGYFVLVRPSYVLGGRAMEIINNKDLRDTYMNRAVKASPEHPVLVDSPYIGKE 754

65 Query: 721 VDAISDGKSVLIPGIMEHIERAGVHSDGSAVYPQQLSKQIQTETATYTKRLAIGLNCI 780

```

-744-

|    |                                                                             |  |
|----|-----------------------------------------------------------------------------|--|
|    | VERI DKG VL+PGINSHIERAGVHSGDSMAVYFPQ L\$+ I +TI +YTKRLAIGLNCI               |  |
|    | Sbjct: 755 VDAICDGEVLLAGINSHIERAGVHSGDSMAVYFPQRLQAIITDIYDVKRLAIGLNCI 814    |  |
| 5  | Query: 781 GGNVQVFVINKQVIVIEVNRASRTVPFLSKVTGIPMAQIATKILQCTLKDLYEGDL 840     |  |
|    | GMMN+QFVI EQVYVIEVNRASRTVPFLSKVT IPMAQ+AT+ILG LKDLTE GL                     |  |
|    | Sbjct: 815 GMMNIQFVIEYEQVYVIEVNRASRTVPFLSKVTNIPMAQLNTQMLGNLKDLYEAGL 874     |  |
|    | Query: 841 YFQSPLVHIAKAFVPSPTKLAQVDSILGSPMESTGEVMSDTSLEKALYKAFAPNNSHLS 900  |  |
|    | P +VH+KAPVPSPTKLA+VDSILGSPMESTG MGS D +LEKALYK+FEA H+++                     |  |
| 10 | Sbjct: 875 APTEDMVHIAKAFVPSPTKLAQVDSILGSPMESTGOLAMGSDVTEALYKSFPAKGLPAD 934  |  |
|    | Query: 901 FGQIVFTIADDSKABLSIARRPKAIGYQINATQGTAAVFAHQGLSACLVGKIGDAANDI 960  |  |
|    | +G ++FT+AD+ K E L+LA+ F IGY ++AT GIRA+ E GL V K+ ++                         |  |
| 15 | Sbjct: 935 YGSLVFTVADSKKESTLALAKDFAEIGTSLVATAGTAFLKENGILYREVEKLAGGDEB 994   |  |
|    | Query: 961 FILW---RSHVQVAINVTGIKR---TADKDGQIRSSAISQGVPLPTALDTAKAMLAVL 1014  |  |
|    | TLW R G VQA+VNI+G R T DG IR AI +G+ELFT+LDT A+L V+                           |  |
|    | Sbjct: 995 GTLVEDIQGRVQAVVNIWNGNTRASLTATDGFIRQEAISRGILPLFSLDVAAILKVM 1054   |  |
| 20 | Query: 1015 ESRCTNIEAI 1024                                                 |  |
|    | +SR F + I                                                                   |  |
|    | Sbjct: 1055 QRSFTTKNY 1064                                                  |  |
|    | Identities = 141/389 (36%), Positives = 222/389 (56%), Gaps = 16/389 (4%)   |  |
| 25 | Query: 518 BSILVLSGSPIRIQGVFFVATVHSVKAQKAYEAIINNSNPETVSDPSVSKLYPE 577       |  |
|    | + I+++GGCP I GQ EFDYA + A+++ GYE ++NSNP T+ TD ++D +Y E                      |  |
|    | Sbjct: 8 KKIMTIIGSGPTIIQGAEEFDYAGTEACLAKKEGTEVVLNNSPATIMTDEIADTVYIE 67      |  |
| 30 | Query: 578 PLTFEDWNVLDLQPKGVIVQPGQQTAINLAQALSEAG-----VTILQTVEDLRAE 631      |  |
|    | P+T E V ++ E+P ++ GGQT +N+A L\$+ G V +LGT++ +D+AE                           |  |
|    | Sbjct: 68 PITLEFVSKILRKRPDALLPTLQQTGLNWMELSKTGLIELNVELLQTL\$AIDQAE 127      |  |
|    | Query: 632 DRDLFEKALKEGIPQPGQQTATNEEALAAKIGFPVLVRPSYVLGRAMEIVENKED 691      |  |
|    | DR+L\$++ + + P AT BEA+ A KIG+P+V P++ +GG I + +B+                            |  |
| 35 | Sbjct: 128 DRBLFEKELCBSINEPLCASDIATVEAINIADKIGYPTIVQPAFTMGOTGGGICDTEBE 187  |  |
|    | Query: 692 LREYIRTAVKASPEHPILVDSYTFQ-KBEVDAISD-GKSEVLPGINSHIERAGVHSGDS 749  |  |
|    | LRR + +K SP L++ I G KR E + + D ++ ME+ + GVH+GDS                             |  |
| 40 | Sbjct: 188 LREIVANKLKLSPVTCCLIRBSIAGYKEIEYEMRDSADNAIVVCMENFDPVGVHTGDS 247   |  |
|    | Query: 750 MAVYFPQQLSKQIQETIAEYTKRLAIGLNCIMMIVQFVI--KNEQVYVIEVNRASRTV 807   |  |
|    | + P Q L\$ + + + + + L G NVQ + + + VIEWNR SR+                                |  |
|    | Sbjct: 248 IVFAPSQTLSDNEYQMLRDASINIRALKIGGGCTVQLALDPSYIEYVIEVNRVSRGS 307    |  |
| 45 | Query: 808 PFLSKVTGIPMAQIATKILQCTLKDIL--GYEDGLY---PQSPLVHIAKAFVPSPTKLAQ 861 |  |
|    | SK TG P\$A+++ K +G TL ++ + Y P V K F F K                                    |  |
|    | Sbjct: 308 ALASIGATGPIAKMSAKIATGMLDEIINFTVNTKYAMPEALDYVAKIARFPFDKFN 367     |  |
|    | Query: 862 VDSILGSPMESTGEVMSDTSLEKALYKA 890                                 |  |
|    | D LG +MK+TGEVM ++E++L KA                                                    |  |
| 50 | Sbjct: 368 GDRHLQQTQATGEVMAIGRNTRESILKA 396                                 |  |

An alignment of the GAS and GBS proteins is shown below:

|    |                                                                             |  |
|----|-----------------------------------------------------------------------------|--|
|    | Identities = 777/1025 (75%), Positives = 896/1025 (86%), Gaps = 1/1025 (0%) |  |
| 55 | Query: 35 LSLKESGYQVVLVNSNPATIMTDKIDADKYVIEPTLSEFVTRILKRRPDALLPTLGGQT 94    |  |
|    | L+LKESGY+V+LVNSNPATIMTDK+IADKYVIEP+TLSEF RI+RKRRPD+LPTLGGQT                 |  |
|    | Sbjct: 1 LALKESGYKVLVNSNPATIMTDKIDADKYVIEPTLSEFVNRIIRKRRPDALLPTLGGQT 60     |  |
| 60 | Query: 95 GLNWMALGKNGILIELNVELLQTLKSAIDAEKRLFKQIMEENQIPESRIVNSVER 154       |  |
|    | GLNWMALSK GIL++L+ELLQTLKSAID+AEKRLFKQIM+EL+QIPES IV +V+E                    |  |
|    | Sbjct: 61 GLNWMALSKGILDDLEIELLQTLKSAIDAEKRLFKQIMQELDQIPESIVTIVDE 120        |  |
|    | Query: 155 AIQPAEQIYPLIVRPAPFTLGGTGGGNCNQBLVDITTKGLKLSPTVTCCLIRSIAGFK 214   |  |
|    | A+ FA IGYP+IVRPAPFTLGGTGGG+C++E+L+IT GLKLSPTVTCCLIRSIAGFK                   |  |
| 65 | Sbjct: 121 AVTFARDIGYPIVRPAPFTLGGTGGGICBSREELCEITNGLKLSPTVTCCLIRSIAGFK 180  |  |

|        |        |                                                               |      |
|--------|--------|---------------------------------------------------------------|------|
| Query: | 215    | EIEYEVWRKADNRAALVVCNMENPDPVGIHTGDSIVFAPAQTLSDEVNQLLRDASLDITRA | 274  |
| Sbjct: | 181    | EIEYEVWRKADNRAALVVCNMENPDPVGIHTGDSIVFAPAQTLSDEVNQLLRDASLDITRA | 240  |
| 5      | Query: | LKIEGGCNVQLALDPNSFKYYIVIEVNPVSRSSALASKATGYPIAKLAAKIAVGLTLDEV  | 334  |
| Sbjct: | 241    | LKIEGGCNVQLALDPNSFKYYIVIEVNPVSRSSALASKATGYPIAKLAAKIAVGLTLDEH  | 300  |
| 10     | Query: | INPITKTYAMFEFALDYVVAEMPRFPFDKFSODRKLTQMKATGEVMAIGRNIEBSLL     | 394  |
| Sbjct: | 301    | INPITKTYAMFEFALDYVVAEMPRFPFDKFSODRKLTQMKATGEVMAIGRNIEBSLL     | 360  |
| 15     | Query: | KACRSLEIGVWHIKIADLONVSDVILEKIRKABDRIAPYLAALREHYSIEKLSLTSI     | 454  |
| Sbjct: | 361    | KACRSLEIGVWHIKIADLONVSDVILEKIRKABDRIAPYLAALREHYSIEKLSLTSI     | 420  |
| 20     | Query: | DGFLEDKLRIVELEDDLKSNRLDINILKVKVNIKGFSDKAIASLMOQINEDVNRNRKEAG  | 514  |
| Sbjct: | 421    | DGFLEDKLRIVELEDDLKSNRLDINILKVKVNIKGFSDKAIASLMOQINEDVNRNRKEAG  | 480  |
| 25     | Query: | ILPYTAIVDTCAEFDGATPYFYSTYAVENBSLISKASILVLGGSPIRIGQVDFDYAT     | 574  |
| Sbjct: | 481    | ILPYTAIVDTCAEFDGATPYFYSTYAVENBSLISKASILVLGGSPIRIGQVDFDYAT     | 540  |
| 30     | Query: | VHSVKAIRESGFEAIIMNSNPETVSTDFSQDKLYPEPLTFEDVNRVIDLEKPEGVILQF   | 634  |
| Sbjct: | 541    | VHSVKAIRESGFEAIIMNSNPETVSTDFSQDKLYPEPLTFEDVNRVIDLEKPEGVILQF   | 600  |
| 35     | Query: | GGQTAINLAKDLNKAQVKILGTQLEDLDRARNRIQFEATLQALNIPQPPGPTATTEESAV  | 694  |
| Sbjct: | 601    | GGQTAINLAKDLNKAQVKILGTQLEDLDRARNRIQFEATLQALNIPQPPGPTATTEESAV  | 660  |
| 40     | Query: | NAQKIGFVPLVRPSVYVLGGRAMKIVENEDLRHYITTRVAKSPDHPLIDAYLQIKCE     | 754  |
| Sbjct: | 661    | NAQKIGFVPLVRPSVYVLGGRAMKIVENEDLRHYITTRVAKSPDHPLIDAYLQIKCE     | 720  |
| 45     | Query: | VDAISDQNLIPGIMBIHIERSGVHSGDSMAVYPPQTSETTIETIVDTYKRLAIGLNCI    | 814  |
| Sbjct: | 721    | VDAISDQNLIPGIMBIHIERSGVHSGDSMAVYPPQTSETTIETIVDTYKRLAIGLNCI    | 780  |
| 50     | Query: | GMSHQFVKIDQKVYVIEVNPFRASRTLPFLSKVTHIPMAQVATKVLIGDKNFTYGYDL    | 874  |
| Sbjct: | 781    | GMSHQFVKIDQKVYVIEVNPFRASRTLPFLSKVTHIPMAQVATKVLIGDKNFTYGYDL    | 840  |
| 55     | Query: | YFASDWVHLKAPVFSFTKLAQVDSLPGPMKSTGEVMSDINLQALYKAFAPAAVLMHPD    | 934  |
| Sbjct: | 841    | YFASDWVHLKAPVFSFTKLAQVDSLPGPMKSTGEVMSDINLQALYKAFAPAAVLMHPD    | 900  |
| 60     | Query: | YGNIVFTVDDIDKEALSLAKVYOSIGYRIYATQGTALYFDANGLETVLVWKLGENDRH    | 994  |
| Sbjct: | 901    | YGNIVFTVDDIDKEALSLAKVYOSIGYRIYATQGTALYFDANGLETVLVWKLGENDRH    | 959  |
| 65     | Query: | IPDLINKGIQAVINTVGGNNIDNDHALIRRSATIEQGVPLFTSLDTAHAMFKVLERSAF   | 1054 |
| Sbjct: | 960    | IPDLINKGIQAVINTVGGNNIDNDHALIRRSATIEQGVPLFTSLDTAHAMFKVLERSAF   | 1019 |
| 70     | Query: | TLEFVTRILRKERPDAALPTLGGQTGLNMMALSKNGILRELVNELLSTKLISALDKR     | 129  |
| Sbjct: | 580    | TLEFVTRILRKERPDAALPTLGGQTGLNMMALSKNGILRELVNELLSTKLISALDKR     | 633  |

Query: 1055 TLKVL 1059

++ +

Sbjct: 1020 NIEAI 1024

Identities = 145/387 (37%), Positives = 229/387 (58%), Gaps = 16/387 (4%)

Query: 10 INVIGSGPIVIGQAAEFYSGTQACLSLKEBGYQVVLNENPATIMTKDIADKVIYIEPI

I+V+GSGPI IGQ EFDY+ + +++ GY+ ++NSNP T+ TD ++DK+Y EP+

Sbjct: 520 IIVLGGSPIRIGQVEFDYATVHSVKAIQKAGYRAIIMNSNPETVSTDFSQDKLYPEPL

Query: 130 DLFKQLMEELNQPIRSEISVNSVEEAIQFABQIGYPLIVRPATLGGTGGGNCINQEQLV 189  
 DLF++ +EL P P+ + + REA++ A++IGP+VRP++ LGG + +N+E L  
 5 Subjct: 634 DLFKALKELGIPQPGQQTAINREBALEAAKKIOPVLVRPSYVLGGRAMEIVENKEDLR 693

Query: 190 DITTKGLKLPVTOCLIERSLAGFKIEYEVMDADNALVVCNMENFDPGIHTGDSIV 249  
 + +K SP L++ I G KE E + + D + L+ ME+ + G+H+GDS+  
 Subjct: 694 EYIRTAVKASPEHILVDSYIPG-KECEVDALSD-GKSVLIPGIMEHIERACVHGCSMA 751

10 Query: 250 FAPAQTLSDVENQLLRDASLDIIRALKIBGGCHVQLALDPSFKYTVIEVNPVRSSAL 309  
 P Q LS + + + + + L G NVQ + + YVIEVNER SR+  
 Subjct: 752 VYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNSQVTVIEVNPVRASRTVPF 809

15 Query: 310 ASKATGYPIAKLAALVAGLTLDEVINPITKITYAMFRBALDYVAKMRPFYDKFESGD 369  
 SK TG P+A++A K+ +G TL ++ Y P V K P F P K D  
 Subjct: 810 LSKVTGIPMAQINLKLGLQTLKDL--GYEDGLY----PQSLVHIKAPVFSPTKLQVQD 863

Query: 370 RKLGTQMKATGEVMAIGRNIESLLKA 396  
 LG +MK+TGEVM ++E++L KA  
 20 Subjct: 864 SLIGFPMKSTGEVMSGDSISLEKALYA 890

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 659

- 25 A DNA sequence (GBSx0699) was identified in *S. agalactiae* <SEQ ID 2027> which encodes the amino acid sequence <SEQ ID 2028>. This protein is predicted to be carboxyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2401(Affixative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB89872 GB:A7132624 carboxyl phosphate synthetase small  
 subunit [Lactococcus lactis]  
 Identities = 242/355 (68%), Positives = 305/355 (85%)

40 Query: 2 KRLLLEDQSVFEGEARFADVETSGRIVFSTGMTQYQESITDQSYNGQITFTTYPILIGNY 61  
 KRLL+LEDG++FEGEA GA+++ +GE+VF+TGMTQYQESITDQSYNGQI+TFTYP++GNY  
 Subjct: 3 KRLLLEDGT+FEAGALGANLVD+GELVFNVTGMTQYQESITDQSYNGQITFTTYPILIGNY 62

45 Query: 62 GINRDYESI RPTCKGVVYEWARYPSNWRQMTLDEFLEKLGIPGISGIDTRAITKIIR 121  
 G+NRDYESI PICK VV++E A PSNR QM+ DEFLEK K IPIG+G+DTRA+TKI+R  
 Subjct: 63 GVNRDYESIHPCTKAVVHEAARRPSNWRQMSDFELKSKNIPGTVTRAITKIIR 122

50 Query: 122 KHGTMKACLINBQNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGNKIVLVDFGLKH 181  
 +HGTMK L+ + + LQ +VL +Q+E ST AY SP G+ +V+VDFGLKH  
 Subjct: 123 KHGTMKASLVQARDEVHQMSQLQATVLPINQVETSTATAYPSPTGRKVVVDFGLKH 182

55 Query: 182 SILRELSQRQCHITVVPHTTAAQEILNLNPDGVLNPGNPNQLNALQMIQEIQKIP 241  
 SILRELS+R+C+TVVP+ T+A+ELL + PDGV+L+NGPG+P +P A++MI+E+QKIP  
 Subjct: 183 SILRELSKRBNLTVVPVNTSAKELIEMEPDGVMLANGPGDPTDVFRAIRMIKEFKQKIP 242

60 Query: 242 IFGICGHQLFAKANGAKTYKMTFGHRGPNHVRHLQGTQGVDFTSQNBQYAVSRDPPFA 301  
 IFGIC+GHQLF+ ANGA TYKM FGRGPNHVR + TG++DFTSQNBQYAVS E+ PE  
 Subjct: 243 IFGICGHQLFSLANGATTYKMFGRGPNHVRVATGIDFTSQNBQYAVSRDPPFA 302

Query: 302 LFIITHEINDKTVBGVRHKYYPAPSVQPHPDAAFGPHDTSYLFDEFINIMDDQFQ 356



-747-

L ITH EIND +VEGVHKKY+PAFSVQPHDAARGPHD SYLFD+P+++D+P++  
 Sbjct: 303 LMITVRIINDSVSGVHKKYPAFSVQPHDAARGPHDASYLFDPMILMHPFK 357

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2029> which encodes the amino acid sequence <SEQ ID 2030>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3534 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 265/354 (74%), Positives = 309/354 (86%)

Query: 2 KRLLLEDGVSFEGEAGADVETSQIVFSTGMTGYQSSITDQSYNGQIITFTYPLIGNY 61  
 KRLLLEDG+YFEGEAGAD+V+G+IVF+TGMTGYQSSITDQSYNGQI+TFTYPLIGNY

Sbjct: 3 KRLLLEDGVTTFEGEAGADIDVTGIVFTGMTGYQSSITDQSYNGQIITFTYPLIGNY 62

Query: 62 GINRDDVESIRPTCKGVVYEWAEYPSNRQCMTLDLFLKIKGIPGISGIDTRALTKIR 121  
 GINRDDVYST PTCKGVV+ E + SNKR+CMTLD FLK+KGIPGISGIDTRALTKIR

Sbjct: 63 GINRDDVESISPTCKGVVVSFSLASNRKQMTLDLFLKIKGIPGISGIDTRALTKIR 122

Query: 122 KHGTMKACLINEGSHIEALENLQSVLLNDQI EQVSTKLAYSFGVGNVIVDPLGLKH 181  
 +HGTMK+ +G+SI +L+ +VL+ +EQVSTK AY +PG+GNVIVDPLGLKH

Sbjct: 123 KHGTMKATMADGDSIQHLKQLRATVLPNTIEQVSTKATYFARGIKGVIVDPLGLKH 182

Query: 182 SILRELQSQCHITVPHITTAQELINLNDGVLSNGPQNLQNLQIQIGKIP 241  
 SILREL+Q+Q+ITVPH+ TA+E+L LNDGV+L+NSNGPQNL+ LP AL MI+ +QKIP

Sbjct: 183 SILRELQSQCHITVPHITTAQELINLNDGVLSNGPQNLQNLQIQIGKIP 242

Query: 242 IPGICMGHQLFPAKANGAKTYKMTFGHRGFNHAVERHLQGVQVFTSQNHGVAVSRDPFA 301  
 IPGICMGHQLF+ ANGAKT KMTFGHRGFNHAVER+ TG+DFTSQNHGVAVR P+ +

Sbjct: 243 IPGICMGHQLFSLANGAKTYKMTFGHRGFNHAVERHLQGVQVFTSQNHGVAVSRDPOT 302

Query: 302 LFTHTERINDKTVEGVHKKYPAFSVQPHDAARGPHDTSYLFDLPMIDFQ 355  
 L +THE+INDKTVEGV+H+ +PAFSVQPHDAARGPHD SYLFDL+ MID ++

Sbjct: 303 LMTVTHINDKTVEGVHKKYPAFSVQPHDAARGPHDASYLFDLPMIDSWR 356

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 660

A DNA sequence (GBSx0700) was identified in *S.galactiae* <SEQ ID 2031> which encodes the amino acid sequence <SEQ ID 2032>. This protein is predicted to be aspartate carbamoyltransferase (pyrB). Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3260 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF72727 GB:AF264709 aspartate transcarbamoylase [Enterococcus faecalis]  
 Identities = 197/303 (65%), Positives = 250/303 (82%)

-748-

Query: 5 TQTLSLEHFSVLSLGNQVMSLKRISIVKRNPSNIGFDDKYVSNLFFENSTRTTHKSF 64  
 ++ +SL+H ++ E L++KVM LI+R+ E K+ ++ Y+ +NLFFENSTRTTHKSF  
 Sbjct: 5 SERISLKLHLLTAEALTDREVMGLIRAGSFKQAGMHPESRQYFATNLFFENSTRTTHKSF 64

Query: 65 EMALKLGKLTIEPNADTSVNGKSTLYDTILTMASALGVDVCIHRPDIDYKELIASPN 124  
 E+AE KLG+ LBF A SSV KGETLYDT+LTMGA+G+DV VIRH ++ Y+ ELI S  
 Sbjct: 65 EVAEKKLGLVLEFPAESRSVQKGETLYDTVLTMSAIGVDVIRHKGKENVYDELIGSKT 124

Query: 125 IHSATVNGDGSGQHPSQSLDLVTIYERFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRL 184  
 I +I+NGEDGSGQHPSQ LLDL+TIYERFG F+GLK+AI+VD+THSRVAKSNMQ+L RL  
 Sbjct: 125 IQCSIIINGDGSGQHPSQCLLDLMTIYERFGFGLKIAIVGDLTHSRVAKSNMQVLKRL 184

Query: 185 GAEIFFSGPKWYSSQFDEYGGYLPIDQLVDQIDVIMLLRVQHERHDGKGVPSKESYHQ 244  
 GAEI+FGSP+EWY QFD YGY+P+D++V++DV+MLLEVQHERHDK PSKE YH +  
 Sbjct: 185 GAEIYFGSPERWYHQFDYGYVPLDIRVKNVDMMLLRVQHERHDKGKSPSKBGHYLE 244

Query: 245 FGLTKERYKHLRDTAIDMHPAPVNRDVEIASDLVEADKARIVKQMSNGVYARIALEAVL 304  
 +GLT ER L+ AIDMHPAPVNRDVE+ +LVE+ ++RIV QMSNGV+ R+AILR+L  
 Sbjct: 245 YGLTHERATRLQKHAIDMHPAPVNRDVEIASDLVPSLQSRIVQMSNGVYRMAILEAIL 304

Query: 305 NSR 307  
 + + +  
 Sbjct: 305 HGK 307

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2033> which encodes the amino acid sequence <SEQ ID 2034>. Analysis of this protein sequence reveals the following:

Possible site: 38  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 208/300 (69%), Positives = 249/300 (82%)

Query: 8 LSLHFSVLSLELGNQVMSLKRISIVKRNPSNIGFDDKYVSNLFFENSTRTTHKSF 67  
 ++L+ VS+E L+ +E+ LI R E K I + V+NLFFENSTRTTHKSF+EA  
 Sbjct: 26 VALINLVSMEALTYREVGLINRSGEYKAGKVVISDHQDLVANLFFENSTRTTHKSF+VA 85

Query: 68 EMALKLGKLTIEPNADTSVNGKSTLYDTILTMASALGVDVCIHRPDIDYKELIASPN 127  
 E KLG+ L+FNAD S+VNGE+LYDT+LTMASG D+CVIRHP+ DYKEL+ GP I +  
 Sbjct: 86 EKKLGLTVLPNADASAVNGESLYDTVLTMSALGTDICVIRHEDDYKELVESAPITTA 145

Query: 128 IHSATVNGDGSGQHPSQSLDLVTIYERFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRLGAE 187  
 +IVNGDGSGQHPSQ LLDL+TIYERFG F+GLKIAI GDITHSRVAKSNMQ+LHRLGAE  
 Sbjct: 146 IIVNGDGSGQHPSQCLLDLTIYERFGFGLKIAIAGDLTHSRVAKSNMQVLKRLGAE 205

Query: 188 GAEIFFSGPKWYSSQFDEYGGYLPIDQLVDQIDVIMLLRVQHERHDGKGVPSKESYHQ 247  
 ++F GP+EWYS F+ YG Y+ IDQ+ ++DVMMLRVQHERHDGKGVPSKESYHQFGL 247  
 Sbjct: 206 LYFPGPBYWSEAFNAYGTYIAIDQILKELDVMMLRVQHERHDGKGVPSKESYHQFGL 265

Query: 248 TKERYKHLRDTAIDMHPAPVNRDVEIASDLVEADKARIVKQMSNGVYARIALEAVLNSR 307  
 T+ERY+ L+D+AIMHPAPVNRDVEIA LVEA KARIV QN+NGV+ R+AI+EA+LN R  
 Sbjct: 266 TQERYQLKDSAIMHPAPVNRDVEIASDLVEAPKARIVSQNANGVYRMAILEAILNSR 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 661

A DNA sequence (GBSx0701) was identified in *S.galactiae* <SEQ ID 2035> which encodes the amino acid sequence <SEQ ID 2036>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2392 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:MAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]
Identities = 176/422 (41%), Positives = 255/422 (59%), Gaps = 8/422 (1%)

Query: 11 IIKNGLLIDPQSGPNQVSEMLIDQGNKQISKEIDIKGIPIIDASKNIVAPGLDIHVHF 70
 I+KNG +IDP D+L+ GKIK+I K I + IIDA IV PG +DIHVH
Sbjct: 5 IVNGYVIDPQNLGEFDILVNGKIKKIDNHLVPEARIIDANGLVCPGPDIDIRVH 64

Query: 71 REPQGTKEINHTGALSAAVGGFTIVLYMANTHTSSPEIVKQVKEAAKEI-KIETV 129
 R+PGGTKE+I +G+ A GGFTI+ N NTP I + +V + + + + +
Sbjct: 65 RDPGQTKEDIESGRCAVAGGFTIVCMPTNPPIDNTIVNYILQSKSVGLCRVLP 124

Query: 130 ATITKSLNGKDLVNFEELEAGVAGFSDGDIPLDITKVLQKAMSLARKHDVVLSEEDP 169
 TITK GK++ +F L EAG F+DDG P+ D+ V+++A+ LA + V + H ED
Sbjct: 125 GTITKGRNGKLIADFYSLKRAAGVAFDIDGSPVNDSSVMKALESLQGLVFMHCEED 164

Query: 190 SLN-GVLGINHIAQKIYHVGASGLARYSMIARDAMIAIYOTAKVHIHLSSESSEVV 248
 L GV INE + + + AE IARD ++A +T VHIQHS+ S+E++
Sbjct: 185 KLAYGV--INIEGVSALLGLSSRAPEAEIQTARDGLAORTGHHVHIQHVSTKLSLEII 242

Query: 249 DFAQKLGANLAEVTPQEFSEKTNELLLTKGNANKINPLRLEKDRQALIDGLSGSVISII 306
 +F ++ G +T EV P H TE +L GANA++NPLR ++DR ALI+G+K G+I
Sbjct: 243 EFFKEKGKIKITCEVNNHLLFTEKEVINGANARVNFPLRKEDRLALIEGVKRGIDCF 302

Query: 309 ASDHAPHHIMEKAADNISQAPSGVTGLETSALGITYLVSTKRLSMIDFLAANTCPAQL 366
 A+DHAPH EK + + A G+ GL+T+L + L +S+ + T NPA++
Sbjct: 303 ATDHAPHQTFEK--ELVEFAMFGIIGLTALPSALE-LYRKGIIISLKKLIEMTNPARI 359

Query: 369 YGFDAGYLRSGPADIVIFDQAEERIIKAEF-ASKSSNSPFGIDKLGKVIHYHTICNGEIV 427
 G D G L+ G PADI IFD +E I+ E SKS N+P G LK+ YTI +G++V
Sbjct: 360 IGVLDGLTKLGSFPADITIFDPNKWILNEETNLSEKSNTPMGKVLGKGVYITIKDKGMV 419

Query: 428 YQ 429
 Y+
Sbjct: 420 YK 421

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2037> which encodes the amino acid sequence <SEQ ID 2038>. Analysis of this protein sequence reveals the following:

```

Possible site: 35

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 76 - 92 (76 - 92)
 INTEGRAL Likelihood = -0.00 Transmembrane 286 - 302 (286 - 302)

----- Final Results -----
 bacterial membrane --- Certainty= 0.132 (Affirmative) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

-750-

|GB:AE000708 dihydroorotase [Aquifex aeolicus] 316 3e-85

&gt;GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]

- 5 Score = 316 bits (801), Expect = 3e-85  
Identities = 177/422 (41%), Positives = 254/422 (59%), Gaps = 8/422 (1%)
- Query: 2 ILIKNGRVMDPKSQRDQVADVLIDGKQIVKLSAIECQRAQVIDASGLIVAPGLVDIRHV 61  
+++KNG V+DP + D+L++ +I KI I RA++IDA GLIV PG +DIRHV  
10 Sbjct: 4 LIVKNGYVIDPSQNLGSGFDILVNGKIKKIDNMLVPEARI IDAGSLIVCPGDIORHV 63
- Query: 62 FRPQGQTHKEDIHTGALAAAGGVTTVMNMNTNPVSDVETLQEVLASAAREKI -HYT 120  
R+PGQT+KEDI +G+ A AGG TT+V M NINE I + + +L + + +  
15 Sbjct: 64 LRDPGQTYKEDIESGRCVAGGPTTIVCMNTNPIDNTTVMNYILQKSKVGLCRVLP 123
- Query: 121 NASVTQAFNGKIVDTFKALLEGAVSPDDGIPLESSKVLKEAFDLANANQTFISLHEED 180  
++T+ GK++ DP +L EAG V+P+DDG P+ S V++A +LA+ I H ED  
15 Sbjct: 124 TGTITTKERKGLADFYSLKEAGCAVPTDQSPVSDGSMRKALELASQLGVPTDHEED 183
- Query: 181 PQL-NGVLGPNFEGIAERHPFGQNTGVAEYSMIARDMIAYDRQANVHQLKSKESVQV 239  
+L GV+ NEG AE IARD ++A HVHIG+S S+++  
20 Sbjct: 184 DKLAYGVI--NBEVSGALLGSSRAPAEIEIQAIDGILAQRTQGHVHIGHVSTKLSLEI 241
- Query: 240 VAFAPQLGAKVTAEVSPQHFSTTEDLLLACTSANGSPPLRTQDRRLAVIEGLKSGVITV 299  
+ F ++ G K+T EV+P H TE +L +G +A+NPPLR + DRLA+IEG+K G-I  
25 Sbjct: 242 IEPFKEKGVKITCEVNPMLLFTEREVLNGANARVNPLRKKEEDRLALIEGVKGIIDC 301
- Query: 300 IATDHAPHHKDEKTVDMTKAPSGMTGLETSLSGLTHLVEPGHILTMSELKMYTINPAL 359  
ATDHAPH EK ++ A G+GL+T+L L L G++L L+E T+NFA  
30 Sbjct: 302 FATDHAPHQTFKEELVEF--AMPGLIGLTALPSAL-ELRYKGIISLAKIENFTINPAR 358
- Query: 360 LYGFAGYLAENGPADLVI PADQERLITENF -ASKASNSPFIKNLEKGVVKYTIADGEV 418  
+ G D G L PAD+ IV +E ++ E SK+ N+P G LKG V YTI DG++  
35 Sbjct: 359 IIGVDLGTALKGSPADITIPDPNKGWILNEETWLSKRNTPLGWKVLEKGIYTIADGEM 418
- Query: 419 VY 420  
VY  
Sbjct: 419 VY 420
- 40 An alignment of the GAS and GBS proteins is shown below:  
Identities = 269/420 (64%), Positives = 338/420 (80%)
- Query: 9 MYIIKNGELIDPQSGPNQVSDMLIDGKQIKSKEIDKGIPIIDASNKIVAPGLVDIHV 68  
M +IKNG ++DP+S +QV+D+LID +I +I+ I+ + +IDAS IVAPGLVDIHV  
45 Sbjct: 1 MILIKNGRVMDPKSQRDQVADVLIDGKQIVKLSAIECQRAQVIDASGLIVAPGLVDIHV 60
- Query: 69 HFREPGQTHKEDIHTGALAAAGGVTTVMNMNTNPVSDVETLQEVLASAAREKI -HYT 126  
HFREPGQTHKE+HTGAL+AA GG TT+VMMNTNP IS E +++V SAAKE I I T  
Sbjct: 61 HFREPGQTHKEDIHTGALAAAGGVTTVMNMNTNPVSDVETLQEVLASAAREKI -HYT 120
- Query: 129 VATITKSLNGKDLNVFZELLEAGVAGSPDDGIPLDTQVLRQAMNLRKHVDVLSLHEED 188  
A++T++ NGKD+ P+ LLEAG PSDDGIPL +KVL+EA +LA + +SLHEED  
50 Sbjct: 121 NASVTQAFNGKIVDTFKALLEGAVSPDDGIPLESSKVLKEAFDLANANQTFISLHEED 180
- Query: 189 PSLNGVLGDIENHIAQKIYHVGASGLAERYSMIARDMIAVQTAQKVIHQLSSSEVEVV 248  
P LINGVLG NR LA++ +H CGA+G+ARYSMIARD MIAY QA VRIQHS +ESV+VV  
55 Sbjct: 181 PQLNGVLGPNFEGIAERHPFGQNTGVAEYSMIARDMIAYDRQANVHQLKSKESVQV 240
- Query: 249 DPAQILGANLTAETVTPQFSEKTENLLITKGANAKLNPFLEKDRQALIDGLKSGVISII 308  
PAQ+LGA +TAEV+PQHS TE+LLE G +AK+NPPLR ++DR A+I+GLKSGVI++I  
60 Sbjct: 241 AFAQQLGAKVTAEVSPQHFSTTEDLLLACTSANGSPPLRTQDRRLAVIEGLKSGVITV 300
- Query: 309 ASDHAPHHIMEKAADNISQAPSGMTGLETSLSGLTIVLSTKELSMIDFLAQNCTNPAQL 368  
A+DHAPH EK D+++APSGMTGLETS+LQ+T+LV L+++ L KMT NPA L  
65 Sbjct: 301 ATDHAPHHKDEKTVDMTKAPSGMTGLETSLSGLTHLVEPGHILTMSELKMYTINPAL 360
- Query: 369 YGFAGYLRGCGPADLVI PDQAEERIKAEPASKSSNSPFIQDKLKGVIHYTICNGEIVY 428

-751-

YGFDAAGYL E GPAD+VIF +ER+I EASK+SNSPFPG+KLGK+ YTI +GE+VY  
 Subjct: 361 YGFDAAGYLAENGPADEVIFADKQERLITENFASKASNSPFGKLGKGVKVTIADGEVVY 420

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 662

A DNA sequence (GBSx0702) was identified in *S. agalactiae* <SEQ ID 2039> which encodes the amino acid sequence <SEQ ID 2040>. This protein is predicted to be orotate phosphoribosyltransferase PyrE (pyrE). Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                         |                                        |
|-------------------------|----------------------------------------|
| bacterial cytoplasm --- | Certainty=0.2214 (Affirmative) < succ> |
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ>   |
| bacterial outside ---   | Certainty=0.0000 (Not Clear) < succ>   |

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA095453 GB:AF068902 orotate phosphoribosyltransferase PyrE  
 [Streptococcus pneumoniae]  
 Identities = 152/208 (73%), Positives = 180/208 (86%)

Query: 1 MDLARQIAMELLDIQAVYLRPQQPPTWASGVKSPIYTDNRVTLSPETRTLIENGFKVQI 60  
 M LA+ IA LL IQAVYL+P+PPTWASG+KSPIYTDNRVTI+YPETRTLIENGFKV I  
 Subjct: 1 MTLAKDIASHLLKIQAVYLRPQEPPTWASGVKSPIYTDNRVTLSPETRTLIENGFKVDI 60

Query: 61 QKHFENVDIAGTATAGIPHGAI IADKMLPPFAYIRSKAKDHGVGNQIEGRVYSGQKQVI 120  
 ++ PP V++IAGTATAGIPHGAI IADKMLPPFAYIRSK KDHG GNQIEGRV GQKQV+  
 Subjct: 61 KEAFPEVEVIAGTATAGIPHGAI IADKMLPPFAYIRSKPKDHGAGNQIEGRVLAQKQKQVI 120

Query: 121 IEDLISTGGSVLEAVTAAQSQIEVLGVVAIPTVQLAKAQAPREADIPIVTLTDYNQLI 180  
 +EDLISTGGSVLEAV AA++ +G +VLGVVAIP+YQL KA++ F +A + LVIT++Y++LI  
 Subjct: 121 VEDLISTGGSVLEAVAAKREKADVLGVVAIPSYQLPKADKPFADAGVKIVTLNYSSELI 180

Query: 181 KVAKVNGYITADQLVLLKKFKEDQNMWQ 208  
 +A+ GYIT + L LLL+FKEDQ NMW  
 Subjct: 181 HLAEQEGYITPGLDILLKRFKEDQNMWQ 208

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2041> which encodes the amino acid sequence <SEQ ID 2042>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                         |                                        |
|-------------------------|----------------------------------------|
| bacterial cytoplasm --- | Certainty=0.1612 (Affirmative) < succ> |
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ>   |
| bacterial outside ---   | Certainty=0.0000 (Not Clear) < succ>   |

An alignment of the GAS and GBS proteins is shown below:

Identities = 158/208 (75%), Positives = 179/208 (85%)

Query: 1 MDLARQIAMELLDIQAVYLRPQQPPTWASGVKSPIYTDNRVTLSPETRTLIENGFKVQI 60  
 M LA QIA +LLDI+AVYL+P+ PPTWASG+KSPIYTDNRVTLSP+TR LIENGFKV+ I  
 Subjct: 1 MTLASQIATQLLDIKAVYLKPEPPTWASGVKSPIYTDNRVTLSPETRTLIENGFKVSTI 60

Query: 61 QKHFENVDIAGTATAGIPHGAI IADKMLPPFAYIRSKAKDHGVGNQIEGRVYSGQKQVI 120  
 + HPP V++IAGTATAGIPHGAI IADKMLPPFAYIRSK KDHG GNQIEGRV GQKQVI  
 Subjct: 61 KEAFPEVEVIAGTATAGIPHGAI IADKMLPPFAYIRSKPKDHGAGNQIEGRVLAQKQKQVI 120

-752-

Query: 121 IEDLISTGGSVLKAIVTAQSQGLRVLVVAIFTYQLAKARQAPREADIPIATLTIDYQILI 180  
 IEDLISTGGSVL+A AA +G +VLGVVAIFTY+L KA Q F+HA I L+TL+Y +LI  
 Sbjct: 121 IEDLISTGGSVLMAAAASREGADV LGVVAIFTYQLPKASQNFKAGIKLITLSNYTLEI 180

Query: 181 KVAKVNGYITADQLVLKFKFKEDQNMWQ 208  
 VAK+ GYIT D L LKFKFKEDQ+NMWQ  
 Sbjct: 181 AVAKLQGYITNDGLHLKFKFKEDQVWQ 208

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 663

A DNA sequence (GBSx0703) was identified in *S.galactiae* <SEQ ID 2043> which encodes the amino acid sequence <SEQ ID 2044>. This protein is predicted to be orotidine 5'-phosphate decarboxylase (pyrF).

- 15 Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have an uncleavable N-term signal seq

- Final Results -----
- 20       bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
       bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
       bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9829> which encodes amino acid sequence <SEQ ID 9830> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC95452 GB:AF068902 orotidine-5'-decarboxylase PyrF  
 [Streptococcus pneumoniae]  
 Identities = 149/231 (64%), Positives = 176/231 (75%), Gaps = 1/231 (0%)

30       Query: 19 MLEKCPILALDPSDLASVTTFLSHFPKRELLFVKIGMELYSEGPISIRYIKSLGHRIFL 78  
           M E PIALDF +Y FL FP EE L++K+QHELY+ GP I+ Y+K LGH +FL  
       Sbjct: 1 MRGHRPIALDPPSFSAVKEFLALPFAESLYLKVGMLYYAAGPEIVSYLKGHGSVFL 60

35       Query: 79 DLKLHDIPTNVRSMSVLAKGLIDMTNVHAAGGVENMKAAAREGLGKGPILLAVTQLTSTS 138  
           DLKLHDIPTNV+S+M VL++LG+DMTNVHAAGGVENMKAAAREGLG L+AVTQLTSTS  
       Sbjct: 61 DLKLHDIPTNVKSAMKVLSQLGVDMTNVHAAGGVENMKAAAREGLGSRKLIATVQLTSTS 120

40       Query: 139 QBEQVQDQHINLSVVDVSVCHYAQAQAEAGLDGVVASAQBNQIKQINHEFICLTGPIRP 198  
           + QMQ Q+I S+ +SV HYA+K EAGLDGVV SAQE IK+ TN PICLTGPIRP  
       Sbjct: 121 EAQMSEFQNTQSLQBSVIHYAKKTAEAGLDGVVCSAQEVQVIKQAINPDFICLTGPIRP 180

      Query: 199 PQTNQLDDQKRTMTFPAQARIVGADYIVVGRPTIKENPQAYLEIKERWNR 249  
           + DQKR MTP A +G+DYIVVGRPTI+AE+P AY IK+EW +  
       Sbjct: 181 AGV-AVGDKRVRTPADAYQIGSDYIVVGRPTIQARDPVAAYHAIDKRWQ 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2045> which encodes the amino acid sequence <SEQ ID 2046>. Analysis of this protein sequence reveals the following:

- Possible site: 44  
 >>> Seems to have no N-terminal signal sequence
- 50       ----- Final Results -----
- bacterial cytoplasm --- Certainty=0.1934 (Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
       bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 55

An alignment of the GAS and GBS proteins is shown below:

-753-

Identities = 149/229 (65%), Positives = 180/229 (78%), Gaps = 1/229 (0%)

Query: 19 MLEKCPITALDFSSDLASVTTTLEHFPKEELLFVKIGWELTYSGPSIIRYIKSLGHRIPL 78  
 M E+ FIALDFS FL+ FP EE L+VKIGWELTY++GP I+RYIKSLGH +FL  
 5 Sbjct: 1 MKEERPIITALDFSSFETKALDLFPABEKIVYKIGWELTYAQGPDIIRYIKSLGHNVPL 60

Query: 79 DLKLDHIDPNTVRSSMVLAKLSDIMTVNVAAGGVEMKAAREGLGKGPILLAVTQITSTS 138  
 DLKLDHIDPNTVR++M+VL +L IDM VEAAGGVEM+KAAEBLG+GP L+AVTQITSTS  
 10 Sbjct: 61 DLKLDHIDPNTVRAMAVLKKLIDMATVHAAGGVEMKAAEBLGQGFPLTAVTQITSTS 120

Query: 139 QEQMQVDQHINLSVVDVSCHYIAQKQAGLGDVVASAQEGWIKKQYTHMFIPLTGTGIRP 198  
 ++QM+ DQ+I S+++SV HY++ A +A LGS V SAGE IK T F CLTGTGIRP  
 10 Sbjct: 121 EDQMGQDQNIQTSLLESVLHYSGAQAQQLDQVCSAQEGVEAIKAVTPTGTFCITGTGIRP 180

Query: 199 FQTMQLDDQKKTMTPEQARIVGADYIVVGRDITKRNFPYQAYLEIKKEW 247  
 +N + DQKR MTP QAR +S+DYTVGRDIT+A++P AY IK EW  
 15 Sbjct: 181 KGSN-IGDQKRWMTFQARRIGSDYIVVGRDITQAKDPVAAYQALKAEW 228

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 664

A DNA sequence (GBSx0704) was identified in *S. agalactiae* <SEQ ID 2047> which encodes the amino acid sequence <SEQ ID 2048> in others. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.70 | Transmembrane | 192 - 208 ( 190 - 211) |
| INTEGRAL | Likelihood = -7.64 | Transmembrane | 226 - 242 ( 218 - 250) |
| INTEGRAL | Likelihood = -3.77 | Transmembrane | 388 - 404 ( 378 - 404) |
| INTEGRAL | Likelihood = -3.08 | Transmembrane | 293 - 309 ( 292 - 311) |
| INTEGRAL | Likelihood = -2.87 | Transmembrane | 165 - 181 ( 162 - 182) |
| INTEGRAL | Likelihood = -2.13 | Transmembrane | 267 - 283 ( 267 - 284) |
| INTEGRAL | Likelihood = -0.90 | Transmembrane | 114 - 130 ( 114 - 130) |
| INTEGRAL | Likelihood = -0.75 | Transmembrane | 318 - 334 ( 318 - 334) |
| INTEGRAL | Likelihood = -0.53 | Transmembrane | 140 - 156 ( 140 - 156) |

----- Final Results -----  
 bacterial membrane --- Certainty=0.4482(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>P:BAB03800 GB:AP001507 unknown conserved protein in others  
 [Bacillus halodurans]  
 Identities = 63/243 (25%), Positives = 120/243 (48%)

Query: 5 MSVLIRAGKLLIESCAEYRVEDIMKHFAKALQIMFAYVSSVSSIIASIGNFYKQENAK 64  
 N + + AG+ + + + + GAE YRVH+T+ + AKA Q N + + V + + I S +  
 5 Sbjct: 8 NDICMLAGEIMLINSIAETRYVESTLERAKAGQFNRVHSFVTTTGIFLSPFESBAGD/WQ 67

Query: 65 VCNVDGVTNMLGRLEAVNLSBQIAKQGLVSPSEIVKQLOLIRHQDYSLVLTISYFG 124  
 + D +L ++ VN +SR+ ++ + E + K ++ + +YS Lt +  
 10 Sbjct: 68 NIRDVDMQDINKVILVQVSRFPVNGEIDAAEALTKLQNLAKQPMNTSPLLEITASGVA 127

Query: 125 AGSPSLALSSLSLDFSAVATGLILGYFLNIMESRITHGFLITLGGSSVVALSANILYFS 184  
 G+FS G +L D+ A + G + + +S + F + + + A LL  
 15 Sbjct: 128 GGAFSYLPCGNLFDITLPAFAGFVSMAYVHLQSLKVRFPARFMAPTGAVAILVL 187

Query: 185 GLGERSIIILGALMVVPGALFVNSVREFSQNNFSTGLALIMSAILLICISISAGVAITI 244  
 GLGE+ +I+G LM +VPG N+VR+ + G+ + +SI+ G+G+ I  
 20 Sbjct: 188 GLGENVDQVITIGTLMPLVPGIPLITNAVRDLISGLLGVTRGACRCPVTSLSIATGIALAI 247

Query: 245 EII 247

-754-

++  
 Sbjct: 248 ALL 250

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 665

- A DNA sequence (GBSx0705) was identified in *S.agalactiae* <SEQ ID 2049> which encodes the amino acid sequence <SEQ ID 2050>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

- Possible site: 40  
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 15 bacterial cytoplasm --- Certainty=0.5134 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9353> which encodes amino acid sequence <SEQ ID 9354> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- >GP:CAB12571 GB:Z99108 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]
- 25 Identities = 193/288 (67%), Positives = 231/288 (80%)
- Query: 1 MNDVINIVYHVENQDLVRYSGDYTNFESVYAMKGAQLRAAYERQOKSIADLQDFVNRNKA 60  
 +N VIN+YHVENQ+L RY GEDY F VY +KK QLEAAY++QQ+R+L+DFV RNKA  
 Sbjct: 222 LNSVINLIVYHVENQELTRYVGDIHQFMEVYEVKQQLRAAYKQQQEQVAKLDFVARNKA 281
- 30 Query: 61 RVATRNAMSRQKLDKMDIIEQLAEKPKPSFEKESRTPGRFIPQAKDLQIGYDRALIK 120  
 RV+TRNAMSRQKLDKMD+IEL AEKPKP F FK +RT G+ IF+ KDL IGYD L++  
 Sbjct: 282 RVSTRNAMSRQKLDKMDIELAEKPKPFEPFKPARTSGKLIFETKDLVIGYDSPLGR 341
- 35 Query: 121 PNLN/TFERNQKIAIVGANGIGKTTLLKSLGIIPIPTSGNVERGDFIDIGYFEQVPGGNR 180  
 PNLN ER KTA+ GANGIGKTTLLKSLG I P+ G+VERG+ I GYFEQEV N  
 Sbjct: 342 PNLNEMERGQKIALYGANIGKTTLLKSLGIIPIPTSGNVERGDIHYTYGFEQVETKTN 401
- 40 Query: 181 QTPLEAVNDAPFALNQAEVRAALAROGILTSKHIESQIQVLSGGEQSKVRFCLLNNRNNV 240  
 T +E VW FP+ Q E+RAA A+CGLT+KHIES++ VLSGGE++KVR C L+N E N+  
 Sbjct: 402 NTCIEBVSWEFPSTQYVETIRAPAKOGLTTKHIESRVSVLSGGEKAKVRLCKLINSETNL 461
- 45 Query: 241 LVLDEPTNHLDDVDAKDELKALKAYKGSITLWCHIEPDTYBGMDDVD 288  
 LVLDEPTNHLDD DAK+ELAKRAK YKGSIL++ HEDPT + W+  
 Sbjct: 462 LVLDEPTNHLDDADAKDELKALKAYKGSITLWCHIEPDTYMDIATETYN 509  
 Identities = 56/219 (25%), Positives = 97/219 (43%), Gaps = 44/219 (20%)
- 50 Query: 104 IFQAKDLQIGY-DRALTKPILN/TFERNQKIAIVGANGIGKTTLLKSLGIIPIPTSGNVER 162  
 I KDL G+ DRN+ ++ + + + +GANG GK+T + + G+ P G VE  
 Sbjct: 3 ILSVKDLSHGFDRAIPNNVSPRLKGEHVGLIGANGBKSTFMMNITGKLSRDPGKVEW 62
- Query: 163 GDFIDLYFEQEVPGGNRQTPLEAVNDAPFALNQAE----- 198  
 + +GY +Q ++ + + DAF L E  
 Sbjct: 63 SKNVRVGYLDQHTVLEKCKSIRDVLKADAPHYLFAMEREMNETYNKMGADPDELKLLRE 122
- 55 Query: 199 ---VRALAR-----OGLTSKHIESQIQVLSGGEQSKVRFCLLNNRNN 239  
 ++ AL GL+ +E + LSGG+++KV L+ +  
 Sbjct: 123 VGVYQDALTNNDPFIYVDSKVERIARGILGSDIGLDRDVTLSGGQRTYVLLAKILLKEPE 182



-755-

Query: 240 VLVLDEPTNHLVDADKDELKRAKAYKGSILMVCHPEPDF 278  
 +L+LDRPTN+LD + LKR L+ Y+ + +++ H+ F  
 Sbjct: 183 LLLLDPTNHLDBQHIEWLKRYLQHYENAFVLISHDISF 221

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2051> which encodes the amino acid sequence <SEQ ID 2052>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2794 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 246/294 (83%), Positives = 274/294 (92%), Gaps = 1/294 (0%)

Query: 1 MNDVINIVHYENQDLVRYSGDYTNFESVYMKKQLEAAAYERQQKEIADLQDFVNRNKA 60  
 +NDVINIVHYENQ LVRV+GDY F++VY NK++QLEAAIERQQKEI+LQDFVNRNKA  
 Sbjct: 233 LNDVINIVHYENQSLVRVYTGDDYQFQAVTENKQSQLEAAAYERQQKEIADLQDFVNRNKA 292

Query: 61 RVATNNAMSRQKGLDFNDIIEQLAEKPKPSFEFKESRTGRFIFQKDLQIGYDRALTK 120  
 RVATNNAMSRQKGLDFNDIIEQLAEKPKP+FEFK+RTF RFIFP K+L IGD LTK  
 Sbjct: 293 RVATNNAMSRQKGLDFNDIIEQLAEKPKPFEFKQARTFSRFIFQTNLVIQTDYDLTK 352

Query: 121 -PLN+TFERNQKIAIVGANGIGK+TLKSLIG+IIPISGVNVERGDVFDLGYFQEVPGGN 179  
 PLN+TFERNQKIAIVGANGIGK+TLKSLIG+I P+ G++ GDF+++GYFQEV G N  
 Sbjct: 353 EPLNITTFERNQKIAIVGANGIGKSTLLKSLIGVIEPLBGHIVTGDFLEVGYFQEVTVGN 412

Query: 180 RQTPLEAVNDAPFALNQAEVRAALAROGSLTSKHIESQIQVLSGGEGKRVFCLLMNRNN 239  
 RQTPLE YNDAPFALNQAEVRAALAROGSLTSKHIESQIQVLSGGEG+KRVFCLLMNRNN  
 Sbjct: 413 RQTPLEVNDAPFALNQAEVRAALAROGSLTSKHIESQIQVLSGGEGKRVFCLLMNRNN 472

Query: 240 VLVLDEPTNHLVDADKDELKRAKAYKGSILMVCHPEDPYEGNMDDVDNFGLS 293  
 VL+LDEPTNHLVD+DAK+ELKRAKAYKGSILMVCHPEDPY GW+ D NDF++L+  
 Sbjct: 473 VLLDDEPTNHLVDADKDELKRAKAYKGSILMVCHPEDPYGNVTDWDFSKT 526  
 Identities = 60/218 (27%), Positives = 102/218 (46%), Gaps = 43/218 (19%)

Query: 104 IPQAKDLQIGY-DRALTKPLN+TFERNQKIAIVGANGIGK+TLKSLIG+IIPISGVNFR 162  
 I + K L G+ DRA+ + + + + I +VGANG GK+T + + G + P G VE  
 Sbjct: 15 ILEVQLSHGPGDRAIFENVSPRLLAGRHIGLVANGSEKSTPMSTVIGHQLPDBGKVEW 74

Query: 163 GDFIDLGYFQEVPGGNRQTPLEAVNDAPFALNQAEVR-----AALA----- 204  
 ++ GY+Q + Q+ + + AF L + E R A+  
 Sbjct: 75 SKYVTAGYLDQHTVLESQYVRDVLRTAFORLFTKTRNINEIYASMAADKADIAVLMEV 134

Query: 205 -----RCGLTSKHIESQIQVLSGGEGKRVFCLLMNRNNV 240  
 G+ +ES+ LSGG+++KV L+ + +  
 Sbjct: 135 GELQDRLESRDFYTLDAKIDEVARALGVMDRGMSDVSLSGGQRTKYLLAKLILEKPI 194

Query: 241 LVLDEPTNHLVDADKDELKRAKAYKGSILMVCHPEPDF 278  
 L+LDRPTNHLVD + + LKR L+ Y+ + +++ H+ F  
 Sbjct: 195 LLLDPTNHLDBQHIEWLKRYLQHYENAFVLISHDISF 222

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-756-

**Example 666**

A DNA sequence (GBSx0706) was identified in *S.agalactiae* <SEQ ID 2053> which encodes the amino acid sequence <SEQ ID 2054>. This protein is predicted to be lipoprotein Nlp1 precursor (pstS). Analysis of this protein sequence reveals the following:

```

5 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2637(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:CAB14429 GB:299116 alternate gene name: yzm@-similar to
 phosphate ABC transporter (binding protein) [Bacillus subtilis]
 Identities = 42/62 (67%), Positives = 49/62 (78%)

 Query: 15 SITSVGTALQPLVEAAADEFGKTNLGTINVGQGGGTGLSQVQSGAVQIGNSDLPARE 74
 S+T GS+AA+QPLV AAA++F + N I VQ GSGGTGLSQV GAVQIGNSD+FARE
20 Sbjct: 45 SLTISGSSAMQPLVLAAAEKPMENPDADIQVQAGSGGTGLSQVSEGAVQIGNSDVFARE 104

 Query: 75 KE 76
 KE
25 Sbjct: 105 KE 106

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1695> which encodes the amino acid sequence <SEQ ID 1696>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 24
 >>> May be a lipoprotein

 ----- Final Results -----
35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

 Identities = 63/74 (85%), Positives = 71/74 (95%)

 Query: 3 LSGCANWIDKQGSITSVGTALQPLVEAAADEFGKTNLGTINVGQGGGTGLSQVQSGA 62
 LE C+*WIDKQ+GIT+VGTALQPLVZA ADEFG +NLGKT+NVQGGGTGLSQVQSGA
40 Sbjct: 20 LSGCSWIDKGSITAVGTALQPLVEAADEFGSSNLGKTVNVQGGGTGLSQVQSGA 79

 Query: 63 VQIGNSDLPAREKE 76
 VQIGNSD+FAREK+
45 Sbjct: 80 VQIGNSDVFAREKD 93

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 667**

50 A DNA sequence (GBSx0707) was identified in *S.agalactiae* <SEQ ID 2055> which encodes the amino acid sequence <SEQ ID 2056>. This protein is predicted to be lipoprotein Nlp1 precursor (pstS). Analysis of this protein sequence reveals the following:

```

 Possible site: 60
 >>> Seems to have an uncleavable N-term signal seq
55

```

-757-

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9343> which encodes amino acid sequence <SEQ ID 9344> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14429 GB:Z99116 alternate gene name: yzmB-similar to  
 phosphate ABC transporter (binding protein) [Bacillus subtilis]  
 Identities = 95/184 (51%), Positives = 126/184 (67%), Gaps = 1/184 (0%)

Query: 3 DEQAVAGLAVIVNKKVAVKILTIHQLRDIFAGKIKWKEVGQQLDISINRAASGSR 62  
 DEQAVY G+A VN VK+++ +L+ IF GKIKWKE+GG+D I+++NR SSG+R  
 Sb|ct: 115 DEQAVVVGMAAAVNPDAEVKDISKDELAKKIPTGKIKWKELGKDKQKITLVNRPSSGTR 174

Query: 63 ATFTNTIMKIVATQSQEQSDSNWKSIVSQTGAIISYLAFAVY-DKSVOTLKLNGFAT 121  
 ATP + P + +DS+ VK I++ TGAII YLAF+Y+ D V L +G+ P  
 Sb|ct: 175 ATTFKYALDGAEPREGITEDSNITVKKI IADTPGAIGYLAFSYLTDDKVTALSIDGVKE 234

Query: 122 AKNVTTDWMKLSYEHMYTKENETGLTKEFLDYMKSDKQVSSIVQEHGYSINDMKVVD 181  
 AKNV T + +W-Y+H YTKG TGL KEFLDY+KS+ +Q SIV GYI + DMKV +D  
 Sb|ct: 235 AKNVATGEYPIWAYQHSYTKGEATGLAKEFLDYLSKESDIQKSIIVTDQGYIVPTDMKVTRD 294

Query: 182 AEGK 185  
 A GK  
 Sb|ct: 295 ANGK 298

There is also homology to SEQ ID 1696.

SEQ ID 9344 (GBS659) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 2 & 3; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 5-7; MW 35kDa) and in Figure 178 (lane 11; MW 35kDa).

GBS659-His was purified as shown in Figure 228, lane 6-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 668

A DNA sequence (GBSx0708) was identified in *S.galactiae* <SEQ ID 2057> which encodes the amino acid sequence <SEQ ID 2058>. This protein is predicted to be phosphate transporter permease PstC (pstC-

2). Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -15.50 | Transmembrane | 35 - 51 ( 27 - 61)     |
| INTEGRAL | Likelihood = -7.64  | Transmembrane | 167 - 183 ( 154 - 186) |
| INTEGRAL | Likelihood = -6.37  | Transmembrane | 282 - 298 ( 277 - 302) |
| INTEGRAL | Likelihood = -5.52  | Transmembrane | 85 - 101 ( 81 - 116)   |
| INTEGRAL | Likelihood = -3.24  | Transmembrane | 133 - 149 ( 131 - 155) |

----- Final Results -----

bacterial membrane --- Certainty=0.7198(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-758-

A related GBS nucleic acid sequence <SEQ ID 863> which encodes amino acid sequence <SEQ ID 863> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
SRCFLG: 0
5 McG: Length of UR: 5
 Peak Value of UR: -0.12
 Net Charge of CR: 2
McG: Discrim Score: -16.22
GvH: Signal Score (-7.5): -4.26
10 Possible site: 41
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program count: 5 value: -15.50 threshold: 0.0
15 INTEGRAL Likelihood = -15.50 Transmembrane 29 - 45 (21 - 55)
 INTEGRAL Likelihood = -7.64 Transmembrane 161 - 177 (148 - 180)
 INTEGRAL Likelihood = -6.37 Transmembrane 276 - 292 (271 - 296)
 INTEGRAL Likelihood = -5.52 Transmembrane 79 - 95 (75 - 110)
 INTEGRAL Likelihood = -3.24 Transmembrane 127 - 143 (125 - 149)
20 PERIPHERAL Likelihood = 0.69 205
 modified ALOM score: 3.60
 icml HYPID: 7 CFP: 0.720

*** Reasoning Step: 3

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.7198(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14428 GB:Z99116 alternate gene name: yzmC-similar to
 phosphate ABC transporter (permease) [Bacillus subtilis]
 Identities = 145/303 (47%), Positives = 209/303 (68%), Gaps = 4/303 (1%)

35 Query: 8 KNQELAKKLTSPSKNSRLKFKGKITFLSLALIVFIVAM-ILIFVAQKGLSTFFVDGKVL 66
 +N ++++L S +N +L++ + + ALI+ ++ I IF+ KGL +F V+GV
 Sbjct: 6 ENMSVSRLLISSRQNRQLDEVGRMIVTACALIMIASVAITIFLGVKGLQSLFVNGVSP 65

Query: 67 TDFLNTKQEP--SAKSPGAFFMIAGSFVITLSAIIATPFAIGAANFMTISPQYKSKI 124
 +FL + W P S +G P I GSF VTLSA+IA P I +FMTEI+P +G K+
40 Sbjct: 66 IEFILTSLNWNPITDSDPKYGVLFFIRGSFAVITLSALIAAPLGIAGPIFMTEIAPNWKGV 125

Query: 125 LQPAVELLVGIPSVVYGFIGLQIIVPFVFSI-FGDTGFGILSGVCVLPWMLPTVTFMTV 183
 LQP +ELLVGIPIVVYGFIGL ++VFP+ GTG +L+G VL VMILPT+T ++
45 Sbjct: 126 LQFVIELLVGIPSVVYGFIGLTVLPVFPFAQKSSGTGHSLLAGTIVSLVMILPTTISISA 185

Query: 184 DSLRAVPRHYKASLAMSATRWQTIWRVILNAAKPGIPTAIVQMARAFGEALAIQMVVG 243
 D++ ++P+ +E S A+GATRWQTI +V++ AA P + TR+V GMARAFGEALAIQMV+G
50 Sbjct: 186 DMGASLPSKLRGSGYALGATRWQTIKRVLNVAAPFTLMTAVVLGMARAFGEALAIQMVVG 245

Query: 244 NSAILPISLITPAATLTSLVLMGIGNTVMGVQNNVMSLAIWLLMSLAINVVIKLTIR 303
 N+ +LP S A TLT++T+ +G+T G+V+NN LWS+ LVL+MS F +I+ ++
55 Sbjct: 246 NTRVLPSPPDTAGITLTITLNMGHTTYGSENNVMSLMSGLVLLVMSFILILIRYLS 305

Query: 304 EKG 306
 K
 Sbjct: 306 RRR 308

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1691> which encodes the amino acid sequence <SEQ ID 1692>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -17.25 Transmembrane 29 - 45 (21 - 55)

```

-759-

INTEGRAL Likelihood = -7.22 Transmembrane 162 - 178 ( 154 - 184)  
 INTEGRAL Likelihood = -5.57 Transmembrane 282 - 296 ( 277 - 302)  
 INTEGRAL Likelihood = -5.41 Transmembrane 96 - 112 ( 81 - 116)  
 INTEGRAL Likelihood = -3.08 Transmembrane 133 - 149 ( 131 - 152)

5 ----- Final Results -----  
 bacterial membrane --- Certainty=0.7899 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 10 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 266/311 (85%), Positives = 290/311 (92%), Gaps = 6/311 (1%)

15 Query: 7 MKNQELAKKLSPSKNSRLKSGKGIITFLSALIVFIVAMILIFVAQKGLSTFFVDGVKL 66  
 M+NQELAKKL SPSKNSRLK FG+ ITPL LALIVFIVAMILIFVAQKGLSTFFVD V L  
 Sbjct: 1 MKNQELAKKLSPSKNSRLSTFGRTITFLCLALIVFIVAMILIFVAQKGLSTFFVDGVNKL 60

Query: 67 TDFLNTKWEPSAKS-----FGAFPMIAGSFIVTILSAIITPFAIGAAGVPMTEISPKY 120  
 DFLF +W+PS K+ GA PMI GSF+VTILSAIITPFAIGAAGVPMTEISPKY  
 20 Sbjct: 61 TDFLFGKEWQPSVKNAGGIPYLGALPMITGSLDVTILSAIITPFAIGAAGVPMTEISPKY 120

Query: 121 GSKILQPAVELLVGIPSVVIGFIGLQIIVFPVRSIFGGTGFGLSGVCVLFVMLPTVTF 180  
 G+K+LQPAVELLVGIPSVVIGFIGLQ+IVFP+RSIFGGTGFGLSGVCVLFVMLPTVTF  
 25 Sbjct: 121 GAKLQPAVELLVGIPSVVIGFIGLQIVFPFMRISIFGGTGFGLSGVCVLFVMLPTVTF 180

Query: 181 MTVDLRAVPKHYKEASLWKGATRWQITWRVILNARPGITPAIVFGMARAFGEALAIQM 240  
 MT DSLRAVPKHY+EAS+AMGATRWQITWRV+LNNARPGITPA++FGMARAFGEALAIQM  
 30 Sbjct: 181 MTVDLRAVPKHYREASAMGATRWQITWRVVLNARPGITPAIVFGMARAFGEALAIQM 240

Query: 241 VVGNBAILPTELTTPAATLTSLVLMGIGNVMSTVQBNVLMALVLLMSLAFNVIKIL 300  
 VVGNB++P+ELTTPAATLTSLVLMGIGNVMSTVQBNVLMALVLL+MSLAFN++KL  
 35 Sbjct: 241 VVGNBAMPELTTTPAATLTSLVLMGIGNVMSTVQBNVLMALVLLMSLAFNVLKIL 300

Query: 301 ITREKKNYER 311  
 IT+E K+NYER  
 40 Sbjct: 301 ITKERKNYER 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 669

A DNA sequence (GBSx0709) was identified in *S. agalactiae* <SEQ ID 2059> which encodes the amino acid sequence <SEQ ID 2060>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2469 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 670

A DNA sequence (GBSx0710) was identified in *S. agalactiae* <SEQ ID 2061> which encodes the amino acid sequence <SEQ ID 2062>. This protein is predicted to be probable abc transporter permease protein in soda-comga intergenic reg. Analysis of this protein sequence reveals the following:

```

5 Possible site: 18
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.24 Transmembrane 20 - 36 (19 - 41)
 INTEGRAL Likelihood = -8.28 Transmembrane 66 - 82 (57 - 88)
10 INTEGRAL Likelihood = -6.90 Transmembrane 260 - 276 (258 - 285)
 INTEGRAL Likelihood = -5.47 Transmembrane 109 - 125 (106 - 129)
 INTEGRAL Likelihood = -2.87 Transmembrane 181 - 197 (178 - 198)

 ----- Final Results -----
15 bacterial membrane --- Certainty=0.4694 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:CAB14427 GB:Z99116 alternate gene name: yzmD-similar to
 phosphate ABC transporter (permease) [Bacillus subtilis]
 Identities = 157/294 (53%), Positives = 225/294 (76%)

 Query: 1 MNKKADKLATTILYSIAAIIVTILASLLIFILVRGLPHVWSFLTGKSSSYRAGGGIGI 60
 Sbjct: 1 MNRKITDKLATGMFLCAAIIPAILVGLFSYIINGVSQLSQFITTSSAIAAGGGIRD 60

 Query: 61 QLVNSFFLLIVTLIIISIPLSIGAGIYLSEYAKKGRILNFRVTCIEILSSLPVSVVGLFGY 120
 Sbjct: 61 QLVNSFYLLITMLITIPLVGGVGVMAEYAPNNKIVDFIRTCIEVILSSLPVIVGMFLG 120

30 Query: 121 LIPVVOFYGFSLIISGALALTVFNLPQMTRSVESLQNVHHTOREAGLALGRMETVIY 180
 Sbjct: 121 LMFVNLGTWGYTIIIGGALALTVFNLPVMVRVTEDAIRSVFDLKEASLALGVSRMHTVKT 180

35 Query: 181 VVVPALPSIVTGVVLASGRIFGEAAALIYTAGQSAPALOWSNMNVLSVTSPIISIFQAE 240
 Sbjct: 181 VLIPSAIPIITIGALLASGRVFGEEAALLFTAGLTIPRLNFTWPNFSETSPINIFRPAE 240

 Query: 241 TLAVHIKVNSEGTIPDATQVSAGSAVLLVVLILFNLARSIGKKLBSKLTSS 294
 Sbjct: 241 TLAVHIWVN+G IPDA ++ G + VL++ +L+NL+AR +G ++ KLT++

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1685> which encodes the amino acid sequence <SEQ ID 1686>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.89 Transmembrane 17 - 33 (8 - 40)
 INTEGRAL Likelihood = -10.19 Transmembrane 260 - 276 (257 - 285)
50 INTEGRAL Likelihood = -5.89 Transmembrane 66 - 82 (57 - 87)
 INTEGRAL Likelihood = -5.47 Transmembrane 109 - 125 (106 - 129)
 INTEGRAL Likelihood = -2.02 Transmembrane 181 - 197 (180 - 197)

 ----- Final Results -----
55 bacterial membrane --- Certainty=0.5755 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

60 Identities = 257/294 (87%), Positives = 278/294 (94%)

 Query: 1 MNKKADKLATTILYSIAAIIVTILASLLIFILVRGLPHVWSFLTGKSSSYRAGGGIGI 60

```

-761-

MNAAK DK+AT LY+IA IIV ILASL+++ILVRGLPH+SWSFLTGKSSSYEAGGGIGI  
 Sbjct: 1 MNAAKDKVATGTLATLGIIVAILASLILYLVRGLPHISWSFLTGKSSSYEAGGGIGI 60  
 Query: 61 QLNSFFLLIVTLIIISIPGLGAGIYLSYAKKGLINFPVTCIEILASLPSVVVGLFGY 120  
 5 QLNSFFLLIVTLIIISIPGL GAGIYL+EYAKKG +TNP+KTCIEILASLPSVVVGLFGY  
 Sbjct: 61 QLNSFFLLIVTLIIISIPGLGAGIYLABYAKKGFTNPFIRTCIEILASLPSVVVGLFGY 120  
 Query: 121 LIFVVQPGPSIISGALALIVFNLFQMTKSVNEDGLQNVHHTQREAGLALGSRMETSIVY 180  
 10 LIFVVQPG+YGPSIISGALALIVFNLFQMTK+VNDL +VHHTQREAGLALG+SRMETSIVY  
 Sbjct: 121 LIFVVQPGYGPSIISGALALIVFNLFQMTKRVNEDSLHVVHHTQREAGLALGSRMETSIVY 180  
 Query: 181 VVVPALPSIVTQVLAGSRIPGBAAALIVTQGSAPALDWSNNVLSVTSISIPRQAE 240  
 VV+PALP +VTG+VLAGSRIPGBAAALIVTQGSAPALDWSNNV LSVTSISIPRQ+E  
 Sbjct: 181 VVVPALPGMTGIVLAGSRIPGBAAALIVTQGSAPALDWSNNVLSVTSISIPRQAE 240  
 15 Query: 241 TLAVHIKVNSEGTIPDATQVSAGSAVLLVILLIIFNLARSIGKKLHSLTSS 294  
 TLAVHIKVNSEGTIPDAT VSAGSAVLL+ ILIFN SA ICKKLHSA+T++  
 Sbjct: 241 TLAVHIKVNSEGTIPDATVVSAGSAVLLIFILIFNFAHFIGKKLHSMATA 294

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 671

A DNA sequence (GBSx0711) was identified in *S. agalactiae* <SEQ ID 2063> which encodes the amino acid sequence <SEQ ID 2064>. This protein is predicted to be phosphate ABC transporter, ATP-binding protein (pstB) (pstB-2). Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 30 bacterial cytoplasm --- Certainty=0.4506 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GF:AAB99016 GB:U67544 phosphate specific transport complex  
 component (pstB) [Methanococcus jannaschii]  
 Identities = 154/247 (62%), Positives = 204/247 (82%)  
 Query: 21 LTTKDLHVYGEKRAKSIDMQFERNKITALIGPSGGCKSTYLRSLNMMIDTIDIAVITG 80  
 40 +TK+L++VYGEK+R+ I++ +NKITALIGPSGGCKST+LR LNR+ND I R+ G  
 Sbjct: 6 METKDLMLVYGEKQALFDINLITYENKITALIGPSGGCKSTFLCRNLRLDLINPVRIG 65  
 Query: 81 QDMYEGIDVNAQDINVYEMKHLIGMVFQRPNPFASKIYNITFAYERAGVKDKKFLDEVV 140  
 +++ +G ++ +D++VYE+RK +GMVQ+DNPFA SIY N+ F G+KDK LD++V  
 45 Sbjct: 66 EVLLDKGNITDKDQDVYELRKRKGMVFOKPNPFAMSIYDNVAFGPRINGIKDKKELDKIV 125  
 Query: 141 ETSLEKQALNDQVKDOLKSAFTLSGGQQRLRCIARAIAVKPEILLMDPEASALDPIATM 200  
 E +LK+ALND+VKD+LHK+A +LSGGQQRLRCIARAIAVKPE+ILLMDPE EALDPI+T+  
 Sbjct: 126 EWLAKKALNDVDEKDELEKNALSLSGGQQRLRCIARAIAVKPEVILLMDPEPTSLDPTL 185  
 50 Query: 201 QLEETNPELKNYITIIIVTHNQQAARASDYTAFFYLEDLIEYDKTNINI PQNAQQSTSD 260  
 ++EE M EL K+YTI++VTHNQQA+R SDYTAFF +G LIE+ +T IF N + + T D  
 Sbjct: 186 KIELLMVLEAKDYTTIVVTHNQQASRVSDYTAFFLMGKLEIFGRTQIFLNPQKREITD 245  
 55 Query: 261 YVSGRFG 267  
 Y+SGRFG  
 Sbjct: 246 YISGRFG 252

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1681> which encodes the amino acid  
 60 sequence <SEQ ID 1682>. Analysis of this protein sequence reveals the following:

-762-

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2796(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 242/267 (90%), Positives = 258/267 (95%)

Query: 1 MAEYNWDERHIITPPEENSALITTKDLHVYVYGEKRAIKGIDMQFEKNKITALIGPSGGGKS 60  
 M EYNW+ERHIIITPPE AL TKDLHVYVY KRAINSDIMQFEK+KITALIGPSGGGKS  
 15 Sbjct: 1 MTEYNWNERHIITPPEETLALATKDLHVYVYGAKEKRAIKGIDMQFEKHKITALIGPSGGGKS 60

Query: 61 TYLRSLNRMNDITDIARVTG+I+Y+GIDVW +D+NVYE+RKH+GMVFCRPNPPAKSIYKN 120  
 TYLRSLNRMNDITDIARVTG+I+Y+GIDVW +D+NVYE+RKH+GMVFCRPNPPAKSIYKN  
 15 Sbjct: 61 TYLRSLNRMNDITDIARVTGELLYQGIDVNRKDMNVYSIRHGLWVFCRPNPPAKSIYKN 120

Query: 121 ITFAERAGVYDKKFLDEIVETSLKQAALNDQVKDLHKSAPFTLSGGQQRLCIARAI+V 180  
 ITFA+ERAGVYDKK LDE+VETSLKQAALNDQVKDLHKSAPFTLSGGQQRLCIARAI+V  
 15 Sbjct: 121 ITFAERAGVYDKKVLDEIVETSLKQAALNDQVKDLHKSAPFTLSGGQQRLCIARAI+V 180

Query: 181 KPEILLMDEPASALDPIATMQLEETMFELKKNYTTIIVTHNQQAARASDYTAFFYLGL 240  
 KP+ILLMDEPASALDPIATMQLEETMFELKKNYTTIIVTHNQQAARASDYTAFFYLGL  
 25 Sbjct: 181 KPEILLMDEPASALDPIATMQLEETMFELKKNYTTIIVTHNQQAARASDYTAFFYLGL 240

Query: 241 IEYDKTNIFQNAQCQSTSDYVSGRFG 267  
 IEYDKT NIFQNA+QCST+DYVSG FG  
 30 Sbjct: 241 IEYDKTRNIFQNAQCQSTNDYVSGHFG 267

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 672

35 A DNA sequence (GBSx0712) was identified in *S.galactiae* <SEQ ID 2065> which encodes the amino acid sequence <SEQ ID 2066>. This protein is predicted to be phosphate ABC transporter, ATP-binding protein (pstB-1). Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3806(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9815> which encodes amino acid sequence <SEQ ID 9816> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

50 >GP:CAB14426 GB:Z99116 alternate gene name: yzmS-similar to  
           phosphate ABC transporter (ATP-binding protein)  
           [Bacillus subtilis]  
           Identities = 148/248 (59%), Positives = 189/248 (75%)

Query: 5 ILQVSDLSVYYNKKKALKEVSMQFYPNEITALIGPSGGGKSTLLRAINRMGDLNFEVTLT 64  
 +L+V DLS+YY K+A+ V+MD N +TALIGPSG GKST LR INRM DL P  
 55 Sbjct: 22 VLEVKDLSIYYGQAVHEVHNMIDIEKNVATLIGPSGGGKSTPLRNINRMNDLIPSAE 81

Query: 65 GAVNYNGHNVSPTDITVELRKRIGIMVFCQNPFPMSVFNVVYGLRKGKDKRATLDEA 124



-763-

G ++Y G N+ + V LR+EIGMVFG+PNEFP S++ N+ + L+ G ++KA LDE  
 Sbjct: 82 GEILYBGLAILGGINVVSIRREIGMVFGKPNFPFKSIYARITHALKYAGERKKAVALDEI 141  
 Query: 125 VETSLKASINDEVKDRILHDSALGLSGGQQQRVCARTLATPKPIILLDEPTSLDPIA 184  
 VE SL A++NDEVKDRILH SAL LSGGQQQR+CIARTLA KP ++LDEP SALDPIS  
 Sbjct: 142 VEESLTKAALNDEVKDRILHSSALSLSGGQQQRICARTLATMKPAVILLDEPASALDPISN 201  
 Query: 185 GKIEETLHGLKDOYTMILAVTRSMQASRISDRITGFFLDGNLIEYGNTEKMFAPKHKETE 244  
 KIEE + GLK +Y++++VT +MQQA R+SDRT FFL+G L+EYG T++F +PK ++TE  
 Sbjct: 202 AKIEELITGLKREYSIIIVTRNMQALRVSDRTAFFLNGELVEYQQTQIPTSPKKQKTE 261  
 Query: 245 DYITGKFG 252  
 DYI GKFG  
 Sbjct: 262 DYINGKFG 269

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2067> which encodes the amino acid sequence <SEQ ID 2068>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3590 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identitles = 208/252 (82%), Positives = 235/252 (92%)  
 Query: 1 NTQPIQLQSDLSVYNNKKALKKEVSMDFYVNEKITALIGPSSGSKSTLLRAINRMGDLNPE 60  
 MT+PIQL+ DLEVYNN+KK LE+VE+D YPNEKITALIGPSSGSKSTLLR+INRM DLNPE  
 Sbjct: 2 MTRPIQLQIRDLGVYNNKXTLKDVSLLDLYPNEKITALIGPSSGSKSTLLRINRMNDLNPE 61  
 Query: 61 VILTAGVNNGHNVISPRDTIVLRKEIGMVFGQPNPFMSVFNENVYGLRLGKIDKAT 120  
 VT+TG+++YNGHN+YSPRDTIV+LRKEIGMVFGQPNPFMS++ENVVYGLRLGKI+DK+  
 Sbjct: 62 VTTTSGIVNGHNIYSPRDTIVDLRKEIGMVFGQPNPFMSIYENVYGLRLAGIRDKSI 121  
 Query: 121 LDEAVETSLKASINDEVKDRILHDSALGLSGGQQQRVCARTLATPKPIILLDEPTSLD 180  
 LD AVE+SLKASIN+EVKDRILHDSA+GLSGGQQQRVCAR LAT P+IILLDEPTSLD  
 Sbjct: 122 LDRAVESSLKASINDEVKDRILHDSAVGLSGGQQQRVCARVLATSPRIILLDEPTSLD 181  
 Query: 181 PISAGKIEETLHGLKDOYTMILAVTRSMQASRISDRITGFFLDGNLIEYGNTEKMFAPKHKETE 240  
 PISAGKIEETL LK YT+ +VTRSMQASR+SDRTGFFL+G+L+E G TK MFMNPK  
 Sbjct: 182 PISAGKIEETLILLKKDYTLAIVTRSMQASRLSDRTGFFLDGGLLECGPTKAMFNPKR 241  
 Query: 241 KETEDYITGKFG 252  
 KETEDIY+GKFG  
 Sbjct: 242 KETEDIYISGKFG 253

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 673

A DNA sequence (GBSx0713) was identified in *S.agalactiae* <SEQ ID 2069> which encodes the amino acid sequence <SEQ ID 2070>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1937 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAD22042 GB:AF118229 Ph0U [Streptococcus pneumoniae]
Identities = 75/216 (34%), Positives = 126/216 (57%), Gaps = 1/216 (0%)

5 Query: 2 LRSKFDEELDKLHNPQFYAMGTEAIGQIKKTVPFVSHDRSLAKEVIEDVTLNPFETKL 61
 +R++FD EL +L F +G + K + A S D+E+A+ +I D +N ++ +E
 Sbjct: 1 MRNPFDEELHLEESQFLGQVLSTASKALALASKDKRMALITNKDHAINQCSAIE 60

10 Query: 62 KKSLEIIALQOPVSGQLRTVITVLKATSDVERMGDHAANAATIRMGGERIPAVELEI 121
 ++ALQOP DLR VI+++ + SD+RMGDH A +AKA +++K E ++ E ++
 Sbjct: 61 LTRALLALQOPVSGDLRFVITSMSSCSDLRMGDHMAIAKAVLQLK-ENQAPDESQL 119

15 Query: 122 NIMGKAVKNMLEALTYINGDDEKAYEVAAMDEIVDDYFRDIQKRVETIQKHDPVAF 181
 + MGK +ML + L A+ KA +A DE +D Y+ + K ++ ++
 Sbjct: 120 HQMKGLSLSLMADLLVAFPLHQASKAISTIAQDKBQIDQYVALSKEIIGLMKDQETSIPN 179

 Query: 182 AKGYFQVLMHLERIGDYGNKNCBWTIVLTKGTIIEI 217
 +Y ++ ELGR DY NICE +VYL+TG+++LL
 Sbjct: 180 GTQVLYIIGHLERFADYIANNICERIVLTGELVDL 215

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1677> which encodes the amino acid sequence <SEQ ID 1678>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2229 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 174/217 (80%), Positives = 194/217 (89%)

35 Query: 1 MLRSKFDEELDKLHNPQFYAMGTEAIGQIKKTVPFVSHDRSLAKEVIEDVTLNPFETKL 60
 MLR+KF+EELDKLHNPFI+MG+E + QI KTVKAFVSHDRSLAKEVIE+D T+NFETKL
 Sbjct: 1 MLRTKFDEELDKLHNPQFYMGMEVLAQINKTVKAFVSHDRSLAKEVIEDTINPFETKL 60

 Query: 61 EKSLEIIALQOPVSGQLRTVITVLKATSDVERMGDHAANAATIRMGGERIPAVELEI 120
 EKSLEIIALQOPVSG DLR VITVLKA+SD+RMGDHAA++AKATIRMGGERIP VE +
40 Sbjct: 61 EKSLEIIALQOPVSGDLRFVITVLKASDIERMGDHAASIAKATIRMGGERIPVVEEQ 120

 Query: 121 INIMGKAVKNMLEALTYINGDDEKAYEVAAMDEIVDDYFRDIQKRVETIQKHDPVAF 180
 IN MGKAVK M+REAL AYIN DD KAYE+AA DEI+D YFR+IQ + VE I+K PD F
45 Sbjct: 121 INIMGKAVKNMVEALNAYINADDTKAYEIAASDEIIDQYFRNIGQVLAIVEIRKSPDAVF 180

 Query: 181 AKGYFQVLMHLERIGDYGNKNCBWTIVLTKGTIIEI 217
 A KGYFQVLM+LERIGDY +NCRNTVILTKGTIIEI
50 Sbjct: 181 AGKEYFQVLMYLERIGDYANNICERIVLTGELVDL 217

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 674

A DNA sequence (GBS0714) was identified in *S.agalactiae* <SEQ ID 2071> which encodes the amino acid sequence <SEQ ID 2072>. This protein is predicted to be aminopeptidase N. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.2845 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA50785 GB:AJ007700 aminopeptidase N [Streptococcus thermophilus]  
 Identities = 556/847 (65%), Positives = 673/847 (78%), Gaps = 4/847 (0%)

10   Query: 3   TVHFVTFKFPVFNLYNLFIDINRQTKTFSGNVAVSGEALDNNISFHQKGLTIKSVLLDNP 62  
           +V F+ P+PENTNLFIDINR KTF+GNVA++GEA+DN+IS HOK LTI SVLLDN+  
   Sbjct: 4   SVARFIESFIPENTNLYNLFIDINRSEKTFIGNVAITGSAIDNHISLHQKDLITNSVLLDNES 63

15   Query: 63   LDPQLDNEAMHIQLHETGSMVLVFEFSGHITDNNYGMPSYTYTNGIKKEVISTQFES 122  
           L+PQ+D+ NEA HI+L ETG + + EFGS ITDNYG+YPSYTT NG KKE+ISTQFES  
   Sbjct: 64   LNPQKDDANEAFHIEPTEPGLTIFIEFSGRITDNNYGMPSYTYTNGEKKIISTQFES 123

20   Query: 123   HFAREVFPSSIDEPEAKATFDLSLKFDQKGEIGALSNMPEINAEQROETGLNFDTPPMS 182  
           HFARE FP +DEPEAKATFDLSLKFD +EG+ ALSNMPEIN+ R+ETG+VTF+TTP+MS  
   Sbjct: 124   HFAREAFPVDSPEAKATFDLSLKFDAAEGUTALSNMPEINSHLREETGVVTFETTPMS 183

25   Query: 183   SYLLAFALGELHOKTHTKMGTLVGSYATKAHQNLDFSLDIVVRVIEFYEDYFGVRY 242  
           +YLLAF G LHKGT TKMGT VG +AT A N +DF+LDI VKVIEFYEDYF V+YF  
   Sbjct: 184   TYLLAFGFGALHGKTAKTMGTVGVPATVAQAENSVDFAIDIAVRIEFYEDYFQVKY 243

30   Query: 243   IPQSLHVALPDFSAGAMENWGLVITYREYVLLVDNENSSSRQQVALVAHEIAHQWFGNL 302  
           IP S +ALPD SAGAMENWGLVITYREYVLLVDNENSS +SRQQVALVAHE+AHQWFGNL  
   Sbjct: 244   IPLSHLALPDLASAGAMENWGLVITYREYVLLVDNENSSAASRQQVALVAHEIAHQWFGNL 303

35   Query: 303   VTMKWMDLMLNESFARMMEYVSIDYIEPKNHFEDPQTG-GLPLALARDATDGVQSV 361  
           VTMKWMDLMLNESFARMMEYVS++ IEP NIEF F G+P ALARDATDGVQSVH+  
   Sbjct: 304   VTMKWMDLMLNESFARMMEYVSVAIEPNSWIFBQFNKLGVALGRDATDGVQSV 363

40   Query: 362   EVNHPDEINTLFDPAIVYAKGSRIMHMLRMLGDTDFRAGLKITPEKIQVNTGRDLN 421  
           EVNHPDRINTLFD AIVYAKGSRIMHMLRMLGDT FA GLK YFEKHQY NT+GRDLN  
   Sbjct: 364   EVNHPDEINTLFDPAIVYAKGSRIMHMLRMLGDEAFAGLKAIFYFEKHQYNTVGRDLN 423

45   Query: 422   ALSQTSKQVAAPMDGWLBPQGPVMAAKIEDEHLITQKQFFIGEHEDKSRMLWQPLMS 481  
           ALS+ SKQDV++PMD+WLBPQGPVPV++A++ +D LIL+QKQFFIGEHEDK RLM+IPLM+  
   Sbjct: 424   ALSEASGKQVSSPMDTWLBPQGPVPVSAEVDVDTLILSKQKQFFIGEHEDKRLWEIPLM 483

50   Query: 482   HWEIGPEILITETTVIPNFSQALAEKNKENGALRPMTENTARYITNYQQQLLHISDLPL 541  
           HW G+P+ L+EE + IPN+SLA +N NG LR NT NTAHYIT+YQQQLL++I+ D  
   Sbjct: 484   HWNGLPDLTSEERIEIPNYSQLATEN--NGVLRMTANTARYITDYQQQLNLELDFAN 541

55   Query: 542   MNIISKLQIVQERHLLAESGMISYSLIPLVSLLSQRTSYLVNSAIKSVIDGLSLFQED 601  
           +D +SKLIQ+QER LLAESG IST+SL+ L+L+ +E S+L++ A ++ GL P+ SD  
   Sbjct: 542   LDTVSKLIQIQERRLLAESGRISYASLVGLLDLVEKESFLISQAKSQILAGLERFIED 601

60   Query: 602   SQDRFDFKEFVNKLGAFFNRLGFEKKEGEGSDSVMHLSLSLALYSNDEHAIEBAHHI 661  
           ++ E +K V++ +F RLGP+ +EGE D+ EMVR +LS ++D ++ A ++  
   Sbjct: 602   TEAEVHYKALVSRQGFNDFRGLGFAKEGSEDEDMVRQTSYLLRADYQPTVLAAANV 661

65   Query: 662   FKAHENNIAPAIARILLVTHMGHIFRSKRLSHLLLETYSYTTDGNFKHQLASLDGTT 721  
           F+AH+ NI +IPA+IR LVL N+MK S L + Y T D NF+RQL ALS+  
   Sbjct: 662   FOAHKENIESIPASIRGLVLINQMKGNSLSLVEHYINAVYNDENFRFQALQALSYLK 721

70   Query: 722   DSKTLKLLSDWKNDIVKPDOLANSYATFLKNSPTFRSVMWMAQENWENIKATLGDM 781  
           + + L +L K+K++VKPDQL +WY FL SF CE+W+WA+ENWENIKA LGDM  
   Sbjct: 722   NQGELDYVLGQLKDKNVKPKDLYL-YYNFWLKSFAQETVMDWAKENWENIKALGDM 780

75   Query: 782   SFDKFTVYIPSSFKTEERLQYKNFEPQLSDMAISRNLSMGIKESARVLLITKQKEV 841  
           SFD FV P+ FK +ERL+QY FPEQ SD A+ RNI MGIK +ARY LI K+K V  
   Sbjct: 781   SFDSPVNIIPAGIFKNQERLQYIAFFEPQISDKALSRNLSMGIKTAARVLLIEKKEAV 840

Query: 842 INTIKKY 848

-766-

+ +K Y  
 Sbjct: 841 ESALKDY 847

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2073> which encodes the amino acid sequence <SEQ ID 2074>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1098 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

15 Identities = 576/848 (67%), Positives = 692/848 (80%), Gaps = 3/848 (0%)

Query: 1 MKTVEHFVTKFVPENNYFLDINRQTKTFSGNVAVSGEALDNNISWQKGLTIKSVLLDN 60  
 MKTVEH + FVPENNY+FLDINRQTKTF+GNVA++GEALDN+++FHQK L INK+LLDN  
 Sbjct: 21 MKTVEHLETFVPENNYFLDINRQTKTF+GNVAINGEALDNHVAFHQKLLDINSLIDN 80

20 Query: 61 QPLDFQLDEDEAMHIQLHETGSMVLVFEFSGHITDNTMGWPGYPTVNGIKKIVSTQF 120  
 + + +Q+D DNE + ++L ETG M LV EFG ITDNTG+YPTIT NG KKEVISTQF  
 Sbjct: 81 EAVIYQVNDNEVVRVELPFGMHTLVIEFGSITDNTGIIPTSYITNGEKKVISTQF 140

25 Query: 121 ESHFAREVFPSTDEPEAKATFSLSKFDQKEGIALSNMPEINARQOETGLNTFDTTPK 180  
 ESHFARE FF IDEP+AKATFDSL FDQ+ GEIALSNMPE+N ++R+ETGLNTEDTT +  
 Sbjct: 141 ESHFAREAFPCIDEPAKATFDSLTLFDQ+GEIALSNMPEINARQOETGLNTFDTT 200

30 Query: 181 MSSYLLA PALGELHGKTHHTKQTLVGSYATKAHQIAELDFSLDIVRVVIEPYEDYPGVR 240  
 MSSYLLA PALGELHGKTH +K GT VG YAT AH L+ LDFSLDI VRVI FYEDYPGV  
 Sbjct: 201 MSSYLLA PALGELHGKTVESKGGTVGVYATTAHPLSSIDFSLDIVRVVIEPYEDYPGV 260

35 Query: 241 YPIQSLHVALPDFSAGAMENWGLTYREYVLLVDENSVSRQQVALVVAHEIAHQWFG 300  
 YPIQSL++ALPDFS+GAMENWGL+TYRE+VLLVDENS+V SRQQVALV+AHIEIAHQWFG  
 Sbjct: 261 YPIQSLINALPDFSAGAMENWGLTYREIYLLVDENSTVQSRQQVALVIAHEIAHQWFG 320

40 Query: 301 NLVTMKNWDDLMLNESFANMEYVSIIDYIEPKINIFEDPQTGGLPALQKDATDGVQSVH 360  
 NLVTMKNWDDLMLNESFANMEYVSI+ IEP I EDPTGG+PALQKDATDGVQSVH  
 Sbjct: 321 NLVTMKNWDDLMLNESFANMEYVSIIEAIEPSWKIIEEDPQTGGLPALQKDATDGVQSVH 380

45 Query: 361 VEVNHPDEINTLPDPAIVYAKSRLNMHLMRLWGLDTPAGLKTYFEHQYQNTIGRDLW 420  
 VEVNHPDEINTLPDPAIVYAKSRLNMHLMRL+GD DFL GL YFEK+QY+NT+IGRDLW  
 Sbjct: 381 VEVNHPDEINTLPDPAIVYAKSRLNMHLMRLRFIGDRDPAIGLHHYFEKYQYNTIGRDLW 440

50 Query: 421 NALSQTSKGDVAAFMDMWLEQGYVFWAAKIEDELLITQKQFFIHEHEDSKRLMIPIN 480  
 N LS TSGKDVAAFMD+MLEQGYFPV+ A++E D+LIL+QKQFFIG+ E+K RLM IPIN  
 Sbjct: 441 NILSDTSKGDVAAFMDMWLEQGYVFWARLENDQLILSQKQFFIHEHEDSKRLMIPIN 500

55 Query: 481 SNWGEIPETLITEETVVPINFSQLAENKENGALFNTENTAHYITVNYQQQLRHIIISDL 540  
 +NW G+PE LTE +VVPINFSQLA +N+ GALRFN +NTAHYIT+YQG LL+ ++++L  
 Sbjct: 501 TNWHLGPELITEAEAVVPINFSQLAENE--GALRFINNTAHYITVNYQQQLDALVLELA 558

60 Query: 541 LMDNISKLQVQRHLLAESQNSYSSSLPIPLVLLSQTSYLVNSAIKSVIDGLSLFVQE 600  
 +DN S LQ++QRR LLA+SG+ISY+ I+ I++ L SY+V A++ V+ GL F+ E  
 Sbjct: 559 QLNTSALQVQRRRLADSGLSYAEVLDAQLAQDDSKSYVWARVQQVSGKRFIDE 618

65 Query: 601 DSQDFDEPKEFVNKLSAPNFVRIQFEKKRGGGDSSEMVHLSLSLALYSNDSHATEAHH 660  
 S E F V + +FN+ GFEX+ E D+ EMVR ++L ++N+ I+  
 Sbjct: 619 GSLAESKYNRLVITTYIDEPNQHGEKKKAEDESDSEDMVQVALGRMLAENCTIDGLRT 678

70 Query: 661 IFKAHRNINIAIPAIRLLVLTNENKHFESKLSHLLSTYSTTDGNFRLQASLSHT 720  
 IF+A++NINIA+IPAA+R LVL N+MK+FE+ L + ETY TTD N + L A S T  
 Sbjct: 679 IFRAVQINIASIPAAVRLLVLANQKYPFETSLVDIYFETVATTDNNLSLVAFSGT 738

721 TDSKTLKKILSDWKNKDIVKQDLAMSWATFLKNSPTQESVWENAOENWENIKATLGGD 780

-767-

T++++L K+KDI+KPQDL+ WY L SFTQ+ +WEGA+ENW+NIK+ LGGD  
 Sbjct: 739 KQPTTIRILVSLKDKDIIKPQDLSY-WYNALGQSPITQDIWEWARENDWIKSALGGD 797

Query: 782 MSFDKFPVIYPSSFKTEERLEQYKNFFEPQLSDMAISRNIISNGIKRISARVLLITKQKE 840  
 MSFUKFVVIYP+S+FKK + L +YK+FFSP+L DMAISRNI+MSI BI ARV LITK+KE  
 Sbjct: 798 MSFDKFPVIYPASNFKTPAHLAEYKSPFEPKLDMAISRNIITGNETARVALITKSKKA 857

Query: 841 VINTIKKY 848  
 VI + Y  
 Sbjct: 858 VIALSHY 865

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 675

A DNA sequence (GBSx0715) was identified in *S.agalactiae* <SEQ ID 2075> which encodes the amino acid sequence <SEQ ID 2076>. This protein is predicted to be response regulator (trcR). Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2741 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA54465 GB:X77249 response regulator [Streptococcus pneumoniae]  
 Identities = 198/224 (88%), Positives = 213/224 (94%)

Query: 1 MIKILLIEDDLGLSNSVDFLDDFADVMQIFDGEGLYEARSQVYDLILLDLMLPEKNGF 60  
 MIKILL+EDDL LNSVDFLDDFADVMQ+FDGEGLYEARSQVYDLILLDLMLPEKNGF  
 Sbjct: 1 MIKILLVEDDLGLSNSVDFLDDFADVMQVFDGEGLYEARSQVYDLILLDLMLPEKNGF 60

Query: 61 QVLKELREKGITTPVLINTAKES+DDKG GF+LGADNYLTKPFYELKMRIQALLKRSG 120  
 QVLKELREKGITTPVLINTAKES+DDKG GF+LGADNYLTKPFYELKMRIQALLKRSG  
 Sbjct: 61 QVLKELREKGITTPVLINTAKES+DDKGHGFELGADNYLTKPFYELKMRIQALLKRSG 120

Query: 121 KPNDSNLIYGDIRDVMSTNSTFPMQTEVELLIGKFDLLVYFLQNQNVILPKSQIFDRIMG 180  
 KPN+N+L YG+I V++STN+ V T VELLIGKFDLLVYFLQNQNVILPK+QIFDR+WG  
 Sbjct: 121 KPNENTLIYGNIVVKLSNTNTVKEVDTFVELLIGKFDLLVYFLQNQNVILPKIQIFDRIMG 180

Query: 181 FDSDTTISVVEVYVSKVRKKLKGTLFSENQLTILRSVGYLKHVE 224  
 FDSDTTISVVEVYVSKVRKKLKGTLFSENQLTILRSVGY+LK V+  
 Sbjct: 181 FDSDTTISVVEVYVSKVRKKLKGTLFSENQLTILRSVGYLLKDVQ 224

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2077> which encodes the amino acid sequence <SEQ ID 2078>. Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2689 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 180/224 (80%), Positives = 200/224 (88%)

-768-

Query: 1 MIKILLIEDDLSELSNSVFDLDDFADVMQVFDGEESLYEASGVYDILLDLMLPEINGF 60  
 MIKILL+EDDLSELSNS+FDLDDFADVMQ+FDG+ESLYEASG+YDILLDLMLPEINGF  
 Sbjct: 1 MIKILLVEDDLSELSNSVFDLDDFADVMQVFDGEESLYEASGVYDILLDLMLPEINGF 60

5 Query: 61 QVLKELREKGIITPVLMINTAKESIDDKQGGFDLGADGVYLPKFFYLEELKMRIOALLRSG 120  
 QVLKELREK I PVLMINTAKE +DDKG GF+LGADGVYLPKFFYLEELKMRIOALLR+G  
 Sbjct: 61 QVLKELREKDIKIPVLMINTAKESLDDKGSGFSLGADGVYLPKFFYLEELKMRIOALLRSG 120

10 Query: 121 KFDNSLIYGDIVDMSTMTSFVNGQTEVLLGKEFDLLVYPLQNNVLPKSOIFDRIRWG 180  
 KP D ++ +G++ VD++ V VELLGKEFDLLVY LQNNVLPK+QIFDR+WG  
 Sbjct: 121 KFDKNISFGNLVVDLARKEVKVSGKVVELLGKEFDLLVYLLQNNVLPKQIFDRIRWG 180

Query: 181 FSDDTTISVVEVYYSKIRKKIKETCFVNRILQTLRSVGYLKHW 224  
 FSDDTTISVVEVY+SK+RKLLKCT F LQTLRSVGYLK+ E  
 15 Sbjct: 181 FSDDTTISVVEVYISKIRKKIKETCFVNRILQTLRSVGYLKHW 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 676

- 20 A DNA sequence (GBSx0716) was identified in *S. agalactiae* <SEQ ID 2079> which encodes the amino acid sequence <SEQ ID 2080>. This protein is predicted to be histidine kinase. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -9.18 Transmembrane 22 - 38 ( 17 - 46)  
 INTEGRAL Likelihood = -4.94 Transmembrane 182 - 198 ( 178 - 201)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>  
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CA54466 GB:X77249 histidine kinase [Streptococcus pneumoniae]  
 Identities = 218/420 (51%), Positives = 305/420 (71%), Gaps = 4/420 (0%)

Query: 17 SHPIHFPTVFGGIFLVMITVILLQVMRYGVYSDVSLKYISTHKNYINMWSRTAA-- 74  
 S+FI F VF+ IF MT+IILQVM +Y+VD L +S +P+ I + ++  
 Sbjct: 15 SYPIRNFQVFTLIFSTMTLLIQVMHSSLYTSDDKLHGLSENFQVILQAINRATERIK 74

40 Query: 75 -LWNSNIASVKLFGGQTVANTDIIILFTSEEVINYPDAFSNYPQLKPNKNGGISELT 133  
 L+N+ + K++ +N++ILF + + + F +K KK LG I ++  
 Sbjct: 75 DLENARADASKVELKPNVSSNTEVILFDKDPITQLLGNRFLGLDKIKLEKKELGHIYQIQ 134

45 Query: 134 L/TNIFGQDETYHATVTKVN-NPAYPNVTYMTAIVNIDQLVNAKERYEKIIIFVMITFWII 192  
 + N +GQ+E Y + ++ N + N+ Y ++N QL A ++++I+ VM +SWI+  
 Sbjct: 135 VFNYSQGEETIYVILMETNISVSSTNKAVALINTSOLEQASQKHBLQVVMVASWIL 194

50 Query: 193 SIGASITLAKVAQKPIETNRYERQKAFVENASHRLRTPLAVLQNRLETLPFRKPNATILENS 252  
 S+ AS+YLA+ + +P++E+ ++Q++FVENASHRLRTPLAVLQNRLETLPFRK ATI++ S  
 Sbjct: 195 SILASLYLARVSVPILSMDKQCSFVENASHRLRTPLAVLQNRLETLPFRKPEATIDVS 254

Query: 253 ENIASSLDEVNRMIITLNLNLARRDDGIKPELAVIKPTLSPSIFENYDITOGENGKF 312  
 E+IASSL+EVNRMR LTT+LLNLARRDDGIKPELA + + P++ F NY++I EN + F  
 55 Sbjct: 255 ESTIASSLDEVNRMITLTLNLNLARRDDGIKPELAVPTSPFNTFTFNEMIASENRVF 314

Query: 313 TGHNMIDQSFKTDKTLKQMLTILFDNAIKYTDNDGSIPTTISDTKYLFLAIDNPGFI 372  
 N I + TD+ LLKQLMTILFDNA+KYT+ DG IDF IS TD+ L+L + +DNG GI  
 Sbjct: 315 RFENRIRHTIVDQLLQKQMLTILFDNAKYTERDGEIDFLSATDRNLYLLVSDNGIGI 374

60 Query: 373 SSEDKVRIFDRFYVDKARTRQGGFGLSLAQQVNSLRGNTVIDNKPFGSIFKIKL 432

-769-

S EDK +IFDRFYRVDKARTQ+GGFGLGLSLA+QIV++L+G +TV DNKP+G+IF++K+  
 Sbjct: 375 STEDKKKIFDRFYRVDKARTQGGFGLGLSLAQIVDLAKGTIVYDKNPKGTIFEVKI 434

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2081> which encodes the amino acid  
 sequence <SEQ ID 2082>. Analysis of this protein sequence reveals the following:

Possible site: 67  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -11.09 Transmembrane 19 - 35 ( 14 - 44)  
 INTEGRAL Likelihood = -10.24 Transmembrane 185 - 201 ( 182 - 206)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5437(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GF:CA54466 GB: X7249 histidine kinase [Streptococcus pneumoniae]  
 Identities = 223/436 (51%), Positives = 313/436 (71%), Gaps = 5/436 (1%)  
 Query: 2 NKLAKKEILSDNYNHFFPFAVFTGIFVIMTIIILQIMRFGVYSSVDGSLVSVNNASSYA 61  
 +KAK +D++F F VPT IF MT+ILQ+M +Y+SVD L +S+N +  
 Sbjct: 3 SKLKKIWDYADPSYFIRNFGVFTLIPFTMTIIILQVNHSSLYTSVDDELHGLSEMPQAVI 62  
 Query: 62 NRTMARISSFFYFDTENNIKALFDSGDSKLLGTTPAANTDILFSGANGTILNAPDAPSNYQ 121  
 +R + D EN A DL ++ +NT+ILF +T L++F +  
 Sbjct: 63 QLAINRATESIKOLEN----ARADASKVEIKPNVSSNTEVILFDKQDPTQLSGNRFGLD 118  
 Query: 122 NFHLDKRLGSIETTSLSNPFYQGEKYHTITVGVHKNYPA-VAYMAVVMVQLDRANE 180  
 L+K+ LG I + N YGOEE Y I + +I + +Y +N (L++A++  
 Sbjct: 119 KIKLEKKELGHIYQIQVFNYSYQGEITRVILMETNISSVTNIKYAVLINTSQLEASQ 178  
 Query: 181 RYERIIIIWMSVPLISLISLASYLAKMSKPEILSYEYKQMFVENASHEILRTPLAVLQNR 240  
 ++R++I++VM+ P++S+L+G+YLA+ S +P+LES +KQ+ FVENASHEILRTPLAVLQNR  
 Sbjct: 179 KHEQLIVVWNASFWILSLASLYLARVSFRLLESMDKQKQSFVENASHEILRTPLAVLQNR 238  
 Query: 241 LESLFRKPNETILENHEILASGLDEVNMRILITNLNLRDDGIDNCPHLDLDPFNA 300  
 LE+LFRKP TI++ SE +AGSL+EVNMR LTT+LINLRDDGI P+ +T FFN  
 Sbjct: 239 LETLFRKPEATIMDVESIASLESVNRMLTTLNLRDDGIDKPELAEVPTSPFIT 298  
 Query: 301 IFENYELVAKEYGKIFYPQNVNRLMDKALLQLITLFDNAIKYTDKNGHIEIIVKT 360  
 F NYE+ A E ++F F+M++R++ D- LKQL+TLFDNA+KYT++G I+ ++  
 Sbjct: 299 TFINYENIASENNVRFPENRIHRTIVTDQLLQMLMTLFDNAVKYTEDEGIEDPLISA 358  
 Query: 361 TDGNLLISVDINGPITDEKKKIKFDRFYRVDKARTQGGFGLGLSLAQQIVMSLKGNI 420  
 TD+NL + V DNG GI+ E+KKKIFDRFYRVDKARTQ GGFGGLG+LA+QIV +LEG +  
 Sbjct: 359 TDGNLLYVSDNGIGISTEDKKKIKFDRFYRVDKARTQGGFGLGLSLAQIVDLAKGTIV 418  
 Query: 421 TVKDNPKGSIPEVKL 436  
 TVKDN PKG+IFEVK+  
 Sbjct: 419 TVKDNPKGTIFEVKI 434

An alignment of the GAS and GBS proteins is shown below:

Identities = 265/436 (60%), Positives = 334/436 (75%), Gaps = 10/436 (2%)  
 Query: 7 ISKPKKGV-SGS--HFIRHPTVPSGIFLVMVITLQVMKRGVYSSVDGSLKYSTHPPNY 63  
 ++K KK +SD+ HF HFF VF+GIF+MT+ILQ+MR+GVYSSVDSSL +S +Y  
 Sbjct: 1 NKLAKKEILSDNYNHFFPFAVFTGIFVIMTIIILQIMRFGVYSSVDGSLVSVNNASSY 60  
 Query: 64 INMYMRTAAYLDNSNIASVKLPG-----QQTVAITDILFTSEEVINYPDAPSNY 116  
 N M+R +++ ++ +K P+ G ANTIDILF++ +N PDAPSNY  
 Sbjct: 61 ANRTMARISSFFYFDTENNIKALFDSGDSKLLGTTPAANTDILFSGANGTILNAPDAPSNY 120  
 Query: 117 QPLKPKKKNLGSIGELTILNIFQDQETVHATVKNMNPAYPNVNTAIVNIDQLVNAKE 176  
 Q +K+ LG I +L N +GQ+R YH +TV V+ YP V YK A+VN++QL A R

-770-

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65  
 70  
 75  
 80  
 85  
 90  
 95  
 100  
 105  
 110  
 115  
 120  
 125  
 130  
 135  
 140  
 145  
 150  
 155  
 160  
 165  
 170  
 175  
 180  
 185  
 190  
 195  
 200  
 205  
 210  
 215  
 220  
 225  
 230  
 235  
 240  
 245  
 250  
 255  
 260  
 265  
 270  
 275  
 280  
 285  
 290  
 295  
 300  
 305  
 310  
 315  
 320  
 325  
 330  
 335  
 340  
 345  
 350  
 355  
 360  
 365  
 370  
 375  
 380  
 385  
 390  
 395  
 400  
 405  
 410  
 415  
 420  
 425  
 430  
 435  
 440  
 445  
 450  
 455  
 460  
 465  
 470  
 475  
 480  
 485  
 490  
 495  
 500  
 505  
 510  
 515  
 520  
 525  
 530  
 535  
 540  
 545  
 550  
 555  
 560  
 565  
 570  
 575  
 580  
 585  
 590  
 595  
 600  
 605  
 610  
 615  
 620  
 625  
 630  
 635  
 640  
 645  
 650  
 655  
 660  
 665  
 670  
 675  
 680  
 685  
 690  
 695  
 700  
 705  
 710  
 715  
 720  
 725  
 730  
 735  
 740  
 745  
 750  
 755  
 760  
 765  
 770  
 775  
 780  
 785  
 790  
 795  
 800  
 805  
 810  
 815  
 820  
 825  
 830  
 835  
 840  
 845  
 850  
 855  
 860  
 865  
 870  
 875  
 880  
 885  
 890  
 895  
 900  
 905  
 910  
 915  
 920  
 925  
 930  
 935  
 940  
 945  
 950  
 955  
 960  
 965  
 970  
 975  
 980  
 985  
 990  
 995

SEQ ID 2080 (GBS339d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 9; MW 73kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 185 (lane 5; MW 73kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 677

A DNA sequence (GBSx0717) was identified in *S.agalactiae* <SEQ ID 2083> which encodes the amino acid sequence <SEQ ID 2084>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1783 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9813> which encodes amino acid sequence <SEQ ID 9814> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA348049 GB:U88582 YlxM [Streptococcus mitans]  
 Identities = 95/110 (86%), Positives = 103/110 (93%)  
 Query: 1 MELEKTRNMALFEPTAALLTDKQMYIELYADYSLAEIARESGVSRQAVYDNIKRT 60  
 MSLEKTRNMALFEPTAALLTDKQMYIELYADYSLAEIARE VSRQAVYDNIKRT  
 Sbjct: 1 MELEKTRNMALFEPTAALLTDKQMYIELYADYSLAEIAREFVSRQAVYDNIKRT 60  
 Query: 61 KILEYEMKIAMYSYVRSQIFDILRKTYDDAFLQKISILSSSDNRD 110  
 KILE YEMKIAMYSYVRS+IFD I+KKY +D +LQ KISIL++IDNRD  
 Sbjct: 61 KILEYEMKIAMYSYVRSKIFDILMKYENDPYLQNKISILPTIDNRD 110

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2085> which encodes the amino acid sequence <SEQ ID 2086>. Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence



-771-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1767(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 95/110 (86%), Positives = 103/110 (93%)

Query: 1 MEIEKTRNMALFFFYAALLTDKQMNYIELYADDSYSLAIEAESGVSQAVYDNIKRTE 60  
 MEIEKTRNMALFFFYAALLTDKQMNYIELYADDSYSLAIEAESGVSQAVYDNIKRTE  
 Sbjct: 4 MEIEKTRNMALFFFYAALLTDKQMNYIELYADDSYSLAIEAESGVSQAVYDNIKRTE 63

Query: 61 KILEAYEMKLMHYSYIVRSQIFDDILEKTYDDAFLOEKISILSSIDNRD 110  
 KILE YEMKLMHYSYIVRS+IFDD++ Y D +LQEKISIL+SIDNR+  
 Sbjct: 64 KILETYEMKLMHYSYIVRSQIFDDMIARHPHDEYLOEKISILSSIDNRD 113

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 678

A DNA sequence (GBSx0719) was identified in *S. agalactiae* <SEQ ID 2087> which encodes the amino acid sequence <SEQ ID 2088>. This protein is predicted to be signal recognition particle protein (fih). Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.22 Transmembrane 37 - 53 ( 37 - 53)

## ----- Final Results -----

bacterial membrane --- Certainty=0.1086(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAE48050 GB:U88582 Fih [Streptococcus mutans]

Identities = 437/522 (83%), Positives = 484/522 (92%), Gaps = 7/522 (1%)

Query: 1 MAFESUTERLQGVFNKIRGKIKLSEKDVQVETKEIRLALLEADVALPVVKTIKRVKRA 60  
 MAFESUTERLQGVFNKIRGKIKLSEKDVQVETKEIRLALLEADVALPVVKTIKRVKRA  
 Sbjct: 1 MAFESUTERLQGVFNKIRGKIKLSEKDVQVETKEIRLALLEADVALPVVKTIKRVKRA 60

Query: 61 VGEHIDTLDPTQQIVKIVNEELTDLLGARTSRIEKSPIKPTIIMVGLQAGKITTPAGK 120  
 VGEHIDTLDPTQQIVKIVNEELTDLLGARTSRIEKSPIKPTIIMVGLQAGKITTPAGK  
 Sbjct: 61 VGEHIDTLDPTQQIVKIVNEELTDLLGARTSRIEKSPIKPTIIMVGLQAGKITTPAGK 120

Query: 121 LANKLKEKNARPMMIAADYRPAADQLKLTGSGINVPVFDMGTHSAVEIVTKLGEQA 180  
 LANKLKEKNARPMMIAADYRPAADQLK LG QINVPVFDMGTHSAVEIVTKLGEQA  
 Sbjct: 121 LANKLKEKNARPMMIAADYRPAADQLKLTGSGINVPVFDMGTHSAVEIVTKLGEQA 180

Query: 181 KENRNDYVLIDTAGRLQIDATMQRIHDVKALQPNRILLVVDLSMIGQEAANVAREFNRQ 240  
 KENRNDYVLIDTAGRLQID LM EL D+GA+A PNEILLVVDLSMIGQEAANVAREFNRQ  
 Sbjct: 181 KENRNDYVLIDTAGRLQIDKIMTELRDIKALANPNEILLVVDLSMIGQEAANVAREFNRQ 240

Query: 241 LSISSVVLTKIDGDIRGGAALSVEKITGKPIKFTGTGKEKITDIETHFHPORMASKILMGND 300  
 L ++GV+LTKIDGDIRGGAALSVEKITGKPIKFTGTGKEKITDIETHFHPORMASKILMGND  
 Sbjct: 241 LEVTVGLITKIDGDIRGGAALSVEKITGKPIKFTGTGKEKITDIETHFHPORMASKILMGND 300

Query: 301 LLTLIERASQYDEKRSMLAEKMRNSFDNFIDQLQVQNGMSMEDILKMLGMANN 360  
 LLTLIERASQYDEKRSMLAEKMRNSFDNFIDQLQVQNGMSMEDILKMLGMANN  
 Sbjct: 301 LLTLIERASQYDEKRSMLAEKMRNSFDNFIDQLQVQNGMSMEDILKMLGMANN 360

-772-

Query: 361 PAMKPKVDENIARKRAIVSSMTPEERENFDLLNPSRRRIAAGSGNTFPVD/NKFIKDF 420  
 PA+ N +VDE IARKRAIVSSMTPEERENFDLLNPSRRRIAAGSGNTFPV+VNFKIKDF  
 Sbjct: 361 PALANVEVDEGIARKRAIVSSMTPEERENFDLLNPSRRRIAAGSGNTFPV+VNFKIKDF 420

5 Query: 421 NQAKMMQGVMSGDMKMKMKMGIDPNLPMKMGMDGM-DMSLENNMGQGMFDLSSL 479  
 NQAK+MMQGVMSGDMK+MK+MGI+PNNP+ + MD+S LEQNMQO GMPD+S L  
 Sbjct: 421 NQAKMMQGVMSGDMKMKMKMGIDPNLPMKMGMDGM-----NNMDSALENNMGQGMFDLSSL 474

10 Query: 480 GGDMDPSQMGGGGLKGVKGFAPAKQSMKMRMANMKKAKKKRK 521  
 G +MD SQMGGGGLKGVKGF PA KQMK+NA +MKKAKK++K  
 Sbjct: 475 GANMDVSGMGGGGLKGVKGFAPAKQSMKMKMRMANMKKAKKKRK 516

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2089> which encodes the amino acid sequence <SEQ ID 2090>. Analysis of this protein sequence reveals the following:

15 Possible site: 53  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.22 Transmembrane 39 - 55 ( 39 - 55)

20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

25 Identities = 458/522 (87%), Positives = 489/522 (92%), Gaps = 4/522 (0%)

Query: 1 MAFESLTERLQGVFKHIGKKKLEKDVETTKIRIALLEADVALPVVTKFIKHVRERA 60  
 MAFESLTERLQGVFKHIGKKKLEKDVETTKIRIALLEADVALPVVTKFIKHVRERA 62  
 Sbjct: 3 MAFESLTERLQGVFKHIGKKKLEKDVETTKIRIALLEADVALPVVTKFIKHVRERA 62

30 Query: 61 VGHSEIDTLPQGIQVIVNEELTDLLGAETSEIEKSPKIPTIIMWGLQAGKITTFAK 120  
 +GHEIIDTLPQGIQVIVNEELTDLLGAETSEIEKSPKIPTIIMWGLQAGKITTFAK 122  
 Sbjct: 63 IGHSEIDTLPQGIQVIVNEELTDLLGAETSEIEKSPKIPTIIMWGLQAGKITTFAK 122

35 Query: 121 LANKLKEKNARPMIAADIVRPAADQLKTLGSSQINVPVFDMTGTHSAVDIVRGLQA 180  
 LANKLKEKNARPMIAADIVRPAADQLKTLGSSQINVPVFDMTGTHSAVDIVRGLQA 182  
 Sbjct: 123 LANKLKEKNARPMIAADIVRPAADQLKTLGSSQINVPVFDMTGTHSAVDIVRGLQA 182

40 Query: 181 RENRNDYVLIDTAGRLQIDATLMOELHDVKAIAQPNKILLVVDMSMIGQRAANVAEENRQ 240  
 RENRNDYVLIDTAGRLQIDATLMOELHDVKAIAQPNKILLVVDMSMIGQRAANVAEENRQ 242  
 Sbjct: 183 RENRNDYVLIDTAGRLQIDATLMOELHDVKAIAQPNKILLVVDMSMIGQRAANVAEENRQ 242

45 Query: 241 LSIISGVLTIKIDGTGGGAALSVEITGPKIKFTGTGKIDTIDTTPHPRMNASRILMGMD 300  
 LSIISGVLTIKIDGTGGGAALSVEITGPKIKFTGTGKIDTIDTTPHPRMNASRILMGMD 302  
 Sbjct: 243 LSIISGVLTIKIDGTGGGAALSVEITGPKIKFTGTGKIDTIDTTPHPRMNASRILMGMD 302

50 Query: 301 LLTILTERASQEQYDEKRSMEIAEKMRNTFPDNDPFDQLDQVNMGMEDLLKMLPGMAN 360  
 LLTILTERASQEQYDEKRSMEIAEKMRNTFPDNDPFDQLDQVNMGMEDLLKMLPGMAN 362  
 Sbjct: 303 LLTILTERASQEQYDEKRSMEIAEKMRNTFPDNDPFDQLDQVNMGMEDLLKMLPGMAN 362

55 Query: 361 PAMKPKVDENIARKRAIVSSMTPEERENFDLLNPSRRRIAAGSGNTFPVD/NKFIKDF 420  
 PA+ N KVDENIARKRAIVSSMTPEERENFDLLNPSRRRIAAGSGNTFPVD/NKFIKDF 421  
 Sbjct: 363 PALANIKVDENIARKRAIVSSMTPEERENFDLLNPSRRRIAAGSGNTFPVD/NKFIKDF 421

60 Query: 421 NQAKMMQGVMSGDMKMKMKMGIDPNLPMKMGMDGM-DMSLENNMGQGMFDLSSL 479  
 NQAK MMQGVMSGDMKMKMKMGIDPNLPMKMGMDGM-DMSLENNMGQGMFDLSSL 479  
 Sbjct: 422 NQAKMMQGVMSGDMKMKMKMGIDPNLPMKMGMDGM-DMSLENNMGQGMFDLSSL 479

Query: 480 GGDMDPSQMGGGGLKGVKGFAPAKQSMKMRMANMKKAKKKRK 521  
 GGDMDPSQMGGGGLKGVKGFAPAKQSMKMRMANMKKAKKKRK 521  
 Sbjct: 480 GGDMDPSQMGGGGLKGVKGFAPAKQSMKMRMANMKKAKKKRK 521

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-773-

**Example 679**

A DNA sequence (GBSx0721) was identified in *S. agalactiae* <SEQ ID 2091> which encodes the amino acid sequence <SEQ ID 2092>. This protein is predicted to be SatD. Analysis of this protein sequence reveals the following:

```

5 Possible site: 49
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL likelihood = -1.28 Transmembrane 3 - 19 (2 - 19)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.1510 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9811> which encodes amino acid sequence <SEQ ID 9812> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AG28336 GBU88582 SatD [Streptococcus mutans]
Identities = 106/222 (47%), Positives = 162/222 (72%), Gaps = 2/222 (0%)

20 Query: 13 MYLALIGDIINSKQILERETTFQSQFLMTLSDVYGEELISPFITTAGDEPQALLKFSK 72
 +Y+A+IGD+I+SK I R Q+ + L+ +++ Y E L S FTIT GDEPQALL P
 Sbjct: 2 IYIATIGDLSISKAITNRPKSQQLKLNLANQINKKYKELLKSAFTITTTGDEPQALLVFN 61

25 Query: 73 KVPQIIDHIQALKPFVNRFGLTGNIITISINSNESIGADGPAYWHARSAINIHDKNDY 132
 ++FOIID I L KP +RFG+G+G+I+T IN +SIG+DGPAYWHAR+AI++IHDKNDY
 Sbjct: 62 QIFQIIDELALGFKPIQIRFVGSGSILTEINFEQSIGSDGPAYWHARAADYIHDKNDY 121

 Query: 133 GTVOVAICLDDQNLLETLSLISAGDFIKSKWNTNHFQMLEHLILQDNYQSQFOHKL 192
 G+ ++ L+D + + + +N++++A +FIKSKW +++++ L+ Y+E+ F H+E+
 Sbjct: 122 GSNHVLVDLEDTETSSQ--DAILAACEFIKSKWTVTQIEVIDGLQNGIYSEKFSHKRM 179

 Query: 193 AGLNIEPESALTKRLKASGLKIYLTTRTQAADLVKSCCTQK 234
 A+ ++ PG+ KRLK+SGLKLYLR + A LL+ + + K
 Sbjct: 180 AEKLDLPSSFNKRLKSSGLKIYLRNKKVATYLLNLAIRKEK 221

35

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2093> which encodes the amino acid sequence <SEQ ID 2094>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3744 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 94/213 (44%), Positives = 137/213 (64%), Gaps = 3/213 (1%)

50 Query: 14 YIALIGDIINSKQILERETTFQSQFLMTLSDVYGEELISPFITTAGDEPQALLKFSK 73
 Y+ALIGDII SKQ+ +R Q+ + +L+ + + +S + +T GDEPQ L +
 Sbjct: 3 YIALIGDIIQSKQLTDRSKVQRTLAAYLLDLNKTFPAPYIISKLSLTGDEPQGLFQVDT 62

 Query: 74 VFOIIDHIQALKPFVNRFGLTGNIITISINSNESIGADGPAYWHARSAINIHDKNDY 133
 +F +ID I + + +RFG+G G+I+T IN +SIGADGPAYWHAR AI +IH KNDY
 Sbjct: 63 IPHLIDLIHMD--IPIRFVGSGSILTDINFDISIGADGPAYWHAREARIYIHQKNDY 121

55 Query: 134 TVQOVAICLDDQNLLETLSLISAGDFIKSKWNTNHFQMLEHLILQDNYQSQFOHKL 193
 +A L N + INSL+AGD IK+ N + + + + L+ Y+R F Q+L
 Sbjct: 122 NITTLA--LRTHHQQDVVINSLLAAGDAIKNNRASQWEIPDTLLDLGIYIEYVTDQQLG 179

```

Query: 194 QLENIEPSALTQRLKASGLKIIYLRTTQADLL 226  
 + ++ SAL+KRLK+S +KIYLRTR R + L  
 Sbjct: 180 KQLSLSSALSRLKSSHVKIYLRTRQSLNCL 212

A related GBS gene <SEQ ID 8637> and protein <SEQ ID 8638> were also identified. Analysis of this protein sequence reveals the following:

LipoP: Possible site: -1 Crend: 7  
 MoG: Discrim Score: 4.96  
 GVR: Signal Score (-7.5): -5.46  
 Possible site: 49  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 1 value: -1.28 threshold: 0.0  
 INTSGFAL Likelihood = -1.28 Transmembrane 3 - 19 ( 1 - 19)  
 PERIPHERAL Likelihood = 5.99 74  
 modified ALOM score: 0.76  
 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1510(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8638 (GBS338) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 5; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 11; MW 55kDa).

GBS338-GST was purified as shown in Figure 215, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 680

A DNA sequence (GBSx0722) was identified in *S.agalactiae* <SEQ ID 2095> which encodes the amino acid sequence <SEQ ID 2096>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.6082(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 681

A DNA sequence (GBSx0723) was identified in *S.agalactiae* <SEQ ID 2097> which encodes the amino acid sequence <SEQ ID 2098>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have a cleavable N-term signal seq.

-775-

```

INTEGRAL Likelihood = -9.87 Transmembrane 126 - 142 (124 - 154)
INTEGRAL Likelihood = -8.23 Transmembrane 45 - 61 (41 - 66)
INTEGRAL Likelihood = -5.10 Transmembrane 241 - 257 (236 - 257)
INTEGRAL Likelihood = -4.04 Transmembrane 199 - 215 (198 - 218)
INTEGRAL Likelihood = -0.22 Transmembrane 96 - 112 (96 - 112)

----- Final Results -----
bacterial membrane --- Certainty=0.4949 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AGG28337 GB:U8582 SatS [Streptococcus mutans]
Identities = 54/103 (52%), Positives = 70/103 (67%), Gaps = 2/103 (1%)

Query: 1 MISDFLRDNPILITLFCARFLADPQWQSGLADSKSHSWRGLMRHLIVFLPLAALMILI 60
 +IS FL NP+LITL AHFLADPQWQSQ +AD KS +W L RHL+IV LPL L ++I
Sbjct: 6 VISQFLSGNFVLTILLIAHFLADPQWQSQKQDLKSSNWTYLIRHLIIVALEPILLSVVI 65

Query: 61 PETTLNLSIWGSHVIDSIKKLSYPWVEEGHF--QKAAFIID 101
 P + L+ I+ SH++IDS K L ++ ++ F KA F+ID
Sbjct: 66 PHSFVLSLGLIFSHVIDSGKILJANSFYKRSFKTKAVFLID 108

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2099> which encodes the amino acid sequence <SEQ ID 2100>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -7.59 Transmembrane 125 - 141 (120 - 144)
INTEGRAL Likelihood = -6.58 Transmembrane 222 - 238 (215 - 238)
INTEGRAL Likelihood = -5.04 Transmembrane 47 - 63 (45 - 77)
INTEGRAL Likelihood = -4.62 Transmembrane 179 - 195 (178 - 199)
INTEGRAL Likelihood = -0.43 Transmembrane 67 - 83 (67 - 83)

----- Final Results -----
bacterial membrane --- Certainty=0.4036 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 109/256 (42%), Positives = 146/256 (56%), Gaps = 28/256 (10%)

Query: 2 ISDFLRDNPILITLFCARFLADPQWQSGLADSKSHSWRGLMRHLIVFLPLAALMILIP 61
 +S +L P LTL H L+D+Q QSQ +AD K L HL+ V +PL L ++IP
Sbjct: 5 VSRYLQPTFLTLFLCHVLSDYQLQSQQVADLKEKHLYTLGYHLIGVSLPLCLITLIP 64

Query: 62 ETTLNLSIWGSHVIDSIKKL---SYPWVEEGHFQKAAFIIDQLAHYTCIIIVFYHALPT 118
 + L+L + SH +ID +K S W E R+DQ H L
Sbjct: 65 QAWLMSLLVMISHALDWLKPKNANSKWKREW-----IFLLDQCCHIAISSFAGLRLAG 119

Query: 119 YLPPNHWLPIKHFIVIALVFIIITKPINIVFKIFNPKFQAKELSSLITQCKYKIMKERS 178
 PN WL PI ++ L ++ITKP NIVFK+FF K+Q + +
Sbjct: 120 YILPN-WL-PIS-IATMTVFLILITKPTNIVFKLFFIKYQPDQGEKM----- 163

Query: 179 EDHESTIBEGAMIGNLERLIMAILLISGQYAAIGLVFTAKSIARYDKISKSGVFAEYTL 238
 +T GAGA IG LSR+++ + +I GQ+A+IGLVFTAKSIARY+KIS+S FAEYTL
Sbjct: 164 ---DTIIGAGATIGILRIVIGVCMIMGQFASIGLVFTAKSIARYNKISBPAPAEYTL 219

Query: 239 IGSLSFISIVSLITHWL 254
 IGSLSFIS+SV I W+
Sbjct: 220 IGSLSFISIVFIAAWI 235

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 682

A DNA sequence (GBSx0724) was identified in *S.agalactiae* <SEQ ID 2101> which encodes the amino acid sequence <SEQ ID 2102>. Analysis of this protein sequence reveals the following:

Possible site: 30  
>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP.AAD17886 GB:AP100456 hyaluronate-associated protein precursor  
(Streptococcus equi)  
Identities = 358/521 (68%), Positives = 426/521 (81%), Gaps = 2/521 (0%)

Query: 1 MSSFRKKLFLGSLATLTATTVTLVACNGESNGSDNKV-DNWYIPTEISTLDISKNT 59  
N+ K K L G++ TL A+ L+ACGN+ S D K DMY PTEI TLDISKNT  
Sbjct: 1 NTVLGTACKRLGLAAVTL-ASVAALMAACGNKQASDCKSEINWYTPTEITLTDISKNT 59

Query: 60 DAYSALAIAGNSGNLLRIDKBEKPKPDIAKKVSVSDGLTYTATLRDLNLSGDSKLSAE 119  
D YS LAIGNSGNMLLR D +GK +PDLA+KV VS DGLTYTATLRD LKNSGDS L+AE  
Sbjct: 60 DTSALAIAGNSGNLLRADAKGLQPDIAKKVSVSDGLTYTATLRDLNLSGDSLTAE 119

Query: 120 DFVYTVRRIVDPKTASEYAVLATESHLLNADKINGSDIKDLNKLGVTAENQVTFKLTSP 179  
DFVY+WA+R+VDPKTASEYAVLATESHLL NAG DL+ LGV A GN+V F LT P  
Sbjct: 120 DFVYSGWRVDPKTASEYAVLATESHLLKNAEDINGSDIKDPLDGLGVKADGNVIFLTSP 179

Query: 180 CPQFKYLLAFSNFPMQKQSVBEKVGKDYGTTSIQNIYSGPYLVKDMNGSNKFKLVKNKY 239  
POFK L+PSNF+PQK+S+V+ GHVYGTTS+ QIYSGPY+VKDMNG++G FKLVRNK  
Sbjct: 180 APQFKSLLSFSNFPVQKESFVKDAGHYGTTSKQIYSGPYLVKDMNGTSGTFKLVRNKN 239

Query: 240 YNDSGHVKINSVIVCTIKKFDTAVQMYKQGGIDFAEISGTSATYQANRKNKDVVDASDAR 299  
YND+K+VCT +V VCT+KKFDTAVQMYKQG++DFA ISGTSATY ANK +KDVV +A  
Sbjct: 240 YNDAQNVKTETVNVQVTKKFDTAVQMYKQGLDFAISGTSATYANRKNKDVVPVLEAT 299

Query: 300 TTYLIYNQIGTSVKALTNQKIRQALNLAIDRKGVVKAADVTGSTPAESLVPPKLAFLPGE 359  
T YI+YNQIG++ L + KIRQALNLAIDRKG+V AAVDTGS PA +LVP LAKL +G  
Sbjct: 300 TAXIVYNQIGATBGLANLKIQRQALNLAIDRKGIVSAVDTGSPATALVPTGLAKLSDGT 359

Query: 360 DLSKYTAPQTYNTSKAQLFKBEAEVQGSLSKLTITADSDSPAANAVDVKSTWESA 419  
DL+++ ARGY Y+ +A KLFKBEAE+G+ +L +TTAD+D+PAAK+AVDY+K TWB+A  
Sbjct: 360 DLTEHVARPYKYDKBEAKLFKBEAEARLQKDLATITTTADADAPAAKAVDYIKSTWETA 419

Query: 420 LPGLTVEEKFPVFKQRLDAQENFDFVLFSWGEDYFBGSTPYGLFTINSAYNYGKFSK 479  
LPGLTVEEKFPV FKQRLD KN+NE+V + WGSDFP+GSTPYGLF + SAYNYGKFP+  
Sbjct: 420 LPGLTVEEKFPVFPKQRLDTRKNQFVAVVLWGSDFYFGSTPYGLFGSAYNYGKFTNA 479

Query: 480 EYDNAVQKAITDALKPGDAAANDYKTBARKALEFQDSYINPVNY 520  
+YD AY KA+TDAAL AA+DYK AKKAL+D + YNP+Y  
Sbjct: 480 DYDAAYNKALITDAALMTDAANDYKAAKALYDINALYNPY 520

There is also homology to SEQ ID 318. An alignment of the GAS and GBS proteins is shown below:

Identities = 138/524 (26%), Positives = 222/524 (42%), Gaps = 73/524 (13%)

Query: 7 KKLKFLG-ISLALTLATTVTLVACNGESNGSDN--KVINWYIPTEISTLDISKNTDAYS 63  
KK K+L +S+A L+ + L ACGN++ G K + + +LD +  
Sbjct: 5 KKSGLAAVSVAILSVGA--LAACGNKNGSGSEATKTYKYVFPVNDPKSLDYLITNGSGT 62

-777-

Query: 64 NLAINSGSNNLRIDKRGKPKDLAKKVSYSVDGLTTTATLRLMLQW--SDGSK---LGA 118  
 I LL D+ G P LAK VS DGLTTT LLED + W +DG + ++A  
 Sbjct: 63 IDTVITQMVGDLENDSTGNLVPFLAKDKVSKDGLTTTATLRLDGVSNYTDAGEEYAFVTA 122

5 Query: 119 EDFVYTWRRIVDPKTAASEAYLATSESHLLADKNKISGDIKDLNKLGVIAKNGQ-VTFKLIT 177  
 EDFV + VD K+ + Y E + N + L  
 Sbjct: 123 EDFVTGLKHAVDKSDAILY--VVSDSIRMLKAYKNGEV-DEKEGVKALDDKTVQYTTAN 178

10 Query: 178 SPCTQFKCYLAFSFMFPQKQSVYKVGKDYGTTSKNQI-YSGPYLAKDMKNSNGKFKLVK 236  
 P + +S P +++ GKD+GTT + I +G Y + + S + K  
 Sbjct: 179 KPESYWNKTTYSVLFPVNAKFLKSKGKDFGITDPSSILVNGAYFLSAFT-SKSSMEFHK 237

Query: 237 NKYYWDSKHVKTNISV--IVQTIKKQDTAVQMYKQQIDFAKISQTSATYQ-ANRNKMDV 293  
 N+ YVD+K+V SV P + + + +G+ A + Y+ A KH D +  
 Sbjct: 238 NENYWDKAWGIERVSKLTYSKSDSGSDFYKNEFKGEFVSARLYPNDFYKSAKKYADHI 297

15 Query: 294 D----ASDARTTYIIYN-----QTGSVKALTNQKIRQALMLATDRKG--- 331  
 D R ++ +N Q KAL N+ RGA+ A DR  
 Sbjct: 298 TYQMLTGDIR--HLTWNLNRTSFYQTKKDPACQDAGKGLANHDFRQAQFAPDRASFOA 355

20 Query: 332 -----VVKAAVDTGSTPAESLVPKKIAKL-FNGEDLSKYTAPGYTYFYS 374  
 V V G + S V K+AKL +D++ A YN  
 Sbjct: 356 QTAGQDAKTKALRNMLVPTFTTIGESDPGSEVEKEMAKLGCEWKNVNLADQDGFNYFE 415

25 Query: 375 KAQKLP--KEGLAEVQGS-SLKLTITADSDSPAANKVADVKESTWESALGLTV----- 425  
 KA+ F KE L G + ++L D + A K + E+L V  
 Sbjct: 416 KAKAEPAKAKALTAGVTPFPVQLQVQDQANATVQEAQSFQKQSVASLGKENVIVNL 475

30 Query: 426 EEKFTVFKQR---LEDKJNENFDVLFPSWGDYPEGSTFYGLPT 466  
 E + T + + E + ++D++ WG DY + T+ + +  
 Sbjct: 476 ETETSTHEAQGFYAEPTPEQDYDISSWKGFTYQDPTLYIDMS 519

SEQ ID 2102 (GBS323) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 4; MW 61.3kDa).

- 35 The GBS323-His fusion product was purified (Figure 209, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 306), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 683

A DNA sequence (GBSx0725) was identified in *S. agalactiae* <SEQ ID 2103> which encodes the amino acid sequence <SEQ ID 2104>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence  
 45 INTEGRAL Likelihood = -1.54 Transmembrane 199 - 215 ( 198 - 215)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1617(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC17173 GB:AF065141 unknown [Streptococcus mutans]  
 Identities = 304/356 (85%), Positives = 334/356 (93%)  
 55 Query: 1 MKRELLEKIDELKELMFWYVLEYQSKLSVPYSFTTLYEYLKSYRRPLEMLLDGSAVNC 60  
 M+RELLEKIDELKELMFWYVLEYQSKLSVPYSFTTLYEYLKSYRRP EML+DSGV+N  
 Sbjct: 1 MKRELLEKIDELKELMFWYVLEYQSKLTVPSYFTTLYEYLKSYRRPFEMLLDGSVNA 60

-778-

Query: 61 HHIAETLSVLNLTKKDMEAFILYLRERPLANTRQNVQSQTINRTLSALSSLFKYL 120  
 + ++\*1 L LE+L+KEIMR+FIYLRGR LLN ++ GVSQTINRTLSALSSLFKYL 120  
 Sbct: 61 NKADIPLETLEHLTKKDEMFILYLRERPLANTRNKQVQSQTINRTLSALSSLFKYL 120

Query: 121 TEEVENADGEPIFYFRNVMMKVSTKKKKTSLASRAENIKQKLFNGNETIFLEYIDCEYQN 180  
 TEEVENADGEPIFYFRNVMMKVSTKKKKTSLA+RAENIKQKLFNGNET+EFLEY+DCEY+  
 Sbct: 121 TEEVENADGEPIFYFRNVMMKVSTKKKKTSLA+RAENIKQKLFNGNETIFLEYIDCEYEQ 180

Query: 181 KLSKRALAFFNKNKERDLAIITALLASGVRLSEAVNLDLKDINLNMVMDVTRKGGKRS 240  
 KLSKRAL+ F KNRKRDLAITALLASGVRLSEAVNLDLKD+NLN+M+I+VTRKGGK DS  
 Sbct: 181 KLSKRALSSFRKNKERDLAIITALLASGVRLSEAVNLDLKDVLNLMIEVTRKGGKRS 240

Query: 241 VNVSAFAKPYLANLYDIRKNRYKAENQDIALFLSEYRGVFNRIASSVVERMVAKYSQDFK 300  
 VNVSA FAKPYL NI+ IR+ RYKA+ D+A FLSEYRGVFNRI+DASS+EMVAKYSQDFK  
 Sbct: 241 VNVSAFAKPYLNTITIRGRYKAKKTDIAFLSEYRGVFNRIASSVVERMVAKYSQDFK 300

Query: 301 RVTFHKLRLHTLATRLYDRTKSQVLVSHQLGHASTQVTDLYTHIVNDEQGNALDKL 356  
 +RVTFHKLRLHTLATRLYDRTKSQVLVSHQLGHASTQVTDLYTHIVNDEQGNALDKL  
 Sbct: 301 IRVTFHKLRLHTLATRLYDRTKSQVLVSHQLGHASTQVTDLYTHIVNDEQGNALDKL 356

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2105> which encodes the amino acid sequence <SEQ ID 2106>. Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.54 Transmembrane 211 - 227 ( 210 - 227)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1617 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9139> which encodes the amino acid sequence <SEQ ID 9140>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 60  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.54 Transmembrane 199 - 215 ( 198 - 215)  
 ----- Final Results -----  
 bacterial membrane --- Certainty= 0.162 (Affirmative) < succ>  
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 283/356 (79%), Positives = 321/356 (89%)

Query: 1 MKRELLLEKIDELKEIMFWYVLEYQSKLSPVPSFTTLRYLYKEYRRFLEMILLDSGVANC 60  
 M+RELLLEKI+ K IMFWYVL+YYQSKL+VPYSFTTLRYLYKEY+RF +WL+D+ +  
 Sbct: 13 MKRELLLEKIETIYKAIMEFWYVLDYQSKLAVPYSFTTLRYLYKEYRRFELNMDADITQA 72

Query: 61 HHIAETLSVLNLTKKDMEAFILYLRERPLANTRQNVQSQTINRTLSALSSLFKYL 120  
 IA+I+IS LE+LTKKD+EAF+LYLRER PLN + + G+SQITINRTLSALSSLFKYL  
 Sbct: 73 PKIADIDSLTLEHLTKKDEAFVILYLRERPLSINTYSTEKGLSQITINRTLSALSSLFKYL 132

Query: 121 TEEVENADGEPIFYFRNVMMKVSTKKKKTSLASRAENIKQKLFNGNETIFLEYIDCEYQN 180  
 TEEVEN GEPIFYFRNVMMKVSTKKKKTSLASRAENIKQKLFNGNETIFLEYIDCEYQN  
 Sbct: 133 TEEVENADGEPIFYFRNVMMKVSTKKKKTSLASRAENIKQKLFNGNETIFLEYIDCEYEQ 192

Query: 181 KLSKRALAFFNKNKERDLAIITALLASGVRLSEAVNLDLKDINLNMVMDVTRKGGKRS 240  
 KLS RA + F KNRKRDLAITALLASGVRLSEAVNLDLKD+NLN+M+I+V RKGKGRS  
 Sbct: 193 KLENRAKSSFRKNKERDLAIITALLASGVRLSEAVNLDLKDVLNLMIEVTRKGGKRS 252

Query: 241 VNVSAFAKPYLANLYDIRKNRYKAENQDIALFLSEYRGVFNRIASSVVERMVAKYSQDFK 300



-779-

VNVA PAK YL +YL +R+ RYKAE OD+A FL+EYRGVGNR+DASS+EKMV KYS+DFK  
 Sbjct: 253 VNVAGFAKGYLESYLAVRQRKYKAEQDIAPVLTSEYRGVGNRMDASSIERMVGKYSDFK 312

Query: 301 VRVTPHKLRRFTLATRLYDATKSQVIVSHQLGHSASTQVTDLYTHIVNDEQKALDL 356  
 +RVTPHKLRRFTLATRLYDATKSQVIVSHQLGHS+STQVTDLYTHIVNDEQKALDL L  
 Sbjct: 313 IRVTPHKLRRFTLATRLYDATKSQVIVSHQLGHSSTQVTDLYTHIVNDEQKALDL 368

SEQ ID 2104 (GBS420) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 5; MW 68kDa).

GBS420-GST was purified as shown in Figure 219, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 684

A DNA sequence (GBSx0726) was identified in *S. agalactiae* <SEQ ID 2107> which encodes the amino acid sequence <SEQ ID 2108>. This protein is predicted to be a sensor-like histidine kinase in *idh 3*' region. Analysis of this protein sequence reveals the following:

Possible site: 24  
 >>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                    |
|----------|--------------------|---------------|--------------------|
| INTEGRAL | Likelihood = -7.75 | Transmembrane | 10 - 26 ( 8 - 34)  |
| INTEGRAL | Likelihood = -3.93 | Transmembrane | 37 - 53 ( 35 - 54) |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.4100(Affixmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database:

>GP: CAB16001 GB: Z99124 similar to two-component sensor histidine kinase [YxdJ] [Bacillus subtilis]

Identities = 96/320 (30%), Positives = 172/320 (53%), Gaps = 16/320 (5%)

Query: 2 IRQFLREHLIWIYILYIN--MFVLFFISFYLYKLMPLYFNLSLGNVIVLIGISIWQYSRY 59  
 ++ FLR H + +L+++ +PV F+ F H +LF LG+ ++L G +++ +  
 Sbjct: 1 MKLFLRSHAVILLILLFQLGLVFFVYFAGLH-SFSLFYLIGVOLLILAGLAYRYMD 59

Query: 60 RKIKMLHLKYPNSSQDSPELOPSDYAFNIIITGLEA--REAQKVSFTIECTNHALVIM 117  
 R L D + L S + Q+E + QK+ ET + + +  
 Sbjct: 60 RGVYHNLSSGSGGTIDPI-LGSSVFCSELYKQMLIRLOHQLHETAKLADRYVTHQ 118

Query: 118 WSHQMKVPLAIAISLMAQTNHLD--KEVBOQLKIQHYLETLLAFKFRQYDDFRFEAV 175  
 W HQ+K PL+ I+L+ Q +P ++++++ +++ LETLL + + DF+ EAV  
 Sbjct: 119 WVHVKPLTSVINLIQPED-EPVFEQIKKQVBOIRFGLSTLYSSRLDLFRDPRFEAV 177

Query: 176 SLREVVVEIITKSYKVICLSKSL--SIIIEGDNINIKDKMLTALFSLQVLDNAIKYSNPES 233  
 SL E+ + +I+SYK + + + + D+ TD KNL FA+ QV+ NA+KYS +S  
 Sbjct: 178 SLSLLEQSVIQSYKRFPIQYRVKPMVCDHDIYTDKMLKFAIGQVTVNAVYKYSAGKS 237

Query: 234 -----KIIISIGBESTRIQDYGIGILESDIPLRPSDGTGYNGERHQQKATGMGLYMTKEV 268  
 + + + +DYG+GI +DI R+P+ +TG NG Q++TG+GL++ KB+  
 Sbjct: 238 DRLELNVCDEORTVLEKTYGVGIPGSODIKRVEFDPIYTGMRRRPQESTGIGLHVKEI 297

Query: 289 LSSINLSISVDSKINYGTAV 308  
 LN ++ + S GT+V  
 Sbjct: 298 TDKIAHTVDSISSPGGTSV 317

SEQ ID 2108 (GBS421) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 6; MW 63kDa).

-780-

GBS421-GST was purified as shown in Figure 219, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 685

- 5 A DNA sequence (GBSx0727) was identified in *S. agalactiae* <SEQ ID 2111> which encodes the amino acid sequence <SEQ ID 2112>. Analysis of this protein sequence reveals the following:

Possible site: 37  
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1310 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD10258 GB:AF036964 putative response regulator [Lactobacillus  
sakai]

Identities = 94/222 (42%), Positives = 140/222 (62%), Gaps = 8/222 (3%)

20 Query: 7 KIYIVEDCMTIVSLKHLASVYHVSSV--SNFRDVHQEIIAPQPDLLMDITLPTFNGF 64  
+I IVEDD TI +L ++L + + ++ +F + + +P L+L+DI LP ++GF  
Sbjct: 3 EINIIVEDCPTIANLIAENLE-KWQLKALIPDDFTITFDFLTKPHVLVDLNPVYDGF 61

25 Query: 65 YWTAEELKFLTIPIFISSNDENDMVMALMGDDPISKPFSLAVLDAKLTALLARSQQ 124  
YW ++R+ +PIIFISS + MDVM++NMGGDDF++KFS+ VL AK+ A+LR+  
Sbjct: 62 YWCKRIEVSVKVPIIFISSRSTNDMVMNMGGDDFVNRKFSMEVLIAKINALLRTYN 121

30 Query: 125 FIQGE---LTFGGFTLT-REGLSSQDKSVILSPTEKILSILLMHPKQVVSKEISLEKL 180  
++ Q + + G + + G D V L S E K+L L+ Q+VS+E LL L  
Sbjct: 122 FVDQNTDVIENHGLLNLQSGGQVGDVTVDLSINYEKLLQLFLMRHGQIVSRKLLRAL 181

35 Query: 181 WENDSPIDQNTLMNMTRLRKKIVPIGP-DYIETVRGVYLL 221  
W+++ P+D NTL VM+ LRKKI G DYI T G GY++  
Sbjct: 182 WDERFVDENILTVNINLRKKIKKQGLDYIQTQIQGYII 223

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 686

- 40 A DNA sequence (GBSx0728) was identified in *S. agalactiae* <SEQ ID 2113> which encodes the amino acid sequence <SEQ ID 2114>. This protein is predicted to be permease OrfY. Analysis of this protein sequence reveals the following:

Possible site: 37  
>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -11.62 Transmembrane 55 - 71 ( 49 - 75)  
INTEGRAL Likelihood = -10.30 Transmembrane 197 - 213 ( 192 - 218)  
INTEGRAL Likelihood = -9.13 Transmembrane 152 - 168 ( 141 - 172)  
INTEGRAL Likelihood = -8.70 Transmembrane 624 - 640 ( 619 - 645)  
INTEGRAL Likelihood = -8.44 Transmembrane 222 - 238 ( 219 - 250)  
50 INTEGRAL Likelihood = -7.75 Transmembrane 283 - 299 ( 280 - 307)  
INTEGRAL Likelihood = -7.70 Transmembrane 533 - 549 ( 526 - 552)  
INTEGRAL Likelihood = -6.95 Transmembrane 108 - 124 ( 99 - 140)  
INTEGRAL Likelihood = -4.88 Transmembrane 585 - 601 ( 581 - 610)  
INTEGRAL Likelihood = -3.82 Transmembrane 25 - 41 ( 21 - 47)

-781-

INTEGRAL Likelihood = -0.48 Transmembrane 602 - 618 ( 602 - 618)

----- Final Results -----

- 5                   bacterial membrane --- Certainty=0.5649 (Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9809> which encodes amino acid sequence <SEQ ID 9810> was also identified.

- 10 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF99695 GB:AF267498 permease OrfY [Streptococcus mutans]  
 Identities = 154/665 (23%), Positives = 299/665 (44%), Gaps = 40/665 (6%)

- 15   Query: 4   MPYLKIAWHNLKHSIDQYIFLLASLLLYSLTCSTLLIMBAVGRDMGTAT---VLFLG 60  
           MP KI++HNL + +P+ + + L+ ++ TA +L G  
   Sbjct: 1   MPLPKISFHNLIWNKSLITLYPALMTIFSGFNVLINFLNPSFYNIPARILIDILIG 60

- Query: 61 VIVLSIPAVVMHYSYNILMKQSSSEFGLYNILGMRQRQVARVASLELPIIYIPLISIGS 120  
           I++S+ ++ Y+ + +R+S G++ +LGM K+Q+ ++ LE ++ G  
   20   Sbjct: 61 FILISLLMLLYGRYANRFISDERNSNMGIPLMGKQQLKIITYLEKLYLFTGTFFGGL 120

- Query: 121 LPSAFFAKFYILIFVNIINHAIHLNLSLSLWPFPIICIVTQIPLTILEVPIVIRHNLHSPL 180  
           +P ++K +L H+I + SL +++ I+ + + R+ S  
   25   Sbjct: 121 IPGFVYSKIFLFRNLIVIGDVRQYSLTAISWLLILTFPIYFIYILSEYLLKQSGIT 180

- Query: 181 SLPRKQOGEKEPKNLILAILALVAIAIAYTHALTSQKAPALAVIY-RFFFAVLVLVAG 239  
           +P K + + K + + + L A + + Y ALTS P + + RF +A LV G  
   30   Sbjct: 181 VIFNSKAKRDNPRKTSVFFVGLFGLFALLMGYHFAITS---PNVITSFRPIYAACLVLG 237

- Query: 240 TYLFYISPMHWLAKRLRQNKHYKYKSEHFVSTSQMIFPMKQNAVGLASITLLAWALVPI 299  
           + + S + L +++ + YY FV ++ + R+ NA+ LA+I + + LV++  
   35   Sbjct: 238 IFCPTSSGVIMLLIVIKKRAIYYNQRFFVVLASFHRIRSNALSATICIPSTATLVSL 297

- Query: 300 ATTYSLSYNTQNVVTGLPFSKVSLSIDNSKGEAKNIPEEKLEKLGKSKKEATINYQMI 359  
           + SLY N+V P+V++ S D E L + + + +T Q  
   40   Sbjct: 298 SVLASLYLAKDNWRLSSPRDVTVL---STTDI-----EPNLMADIATQGHVTLINRQ-- 346

- Query: 360 SMPVQSSELNITSKNVKNVDITTKGFMY-----LITQNDPRLGHOLPKLKNQVAYF 413  
           ++ VQS NI H+ + G M +I+ + P + + +LK+++ +  
   45   Sbjct: 347 NLKVSQSVYGNIGS---HLSVDPNGWANDYQITVISLDSFDSNNTHYRKLQHEILTY 403

- Query: 414 VQKSDRLKINLGNKFDVVNLKEA-YVPETNTYNPLGIIPANNKQI-DNIRKAYLV 471  
           V G + G K VK+K ++ + P I N++I I K L  
   50   Sbjct: 404 VSNKAAPSSYTTNGVKTINVQIKRINFISPLRSMQNFITIDNREIQTILKEILT 463

- Query: 472 YTKNINTFPKTFKAYLDLNSQELINSISKNDIIEVDG--KYVGNISTQSFLEKGYOMFG 529  
           + T Y + ++N D+E ++ N+ ++ +PG2  
   55   Sbjct: 464 WG-----TWAGY-HVKGKMKQKQFDYELETINPQPSANVSIQVKSNNALFG 516

- Query: 530 LLPTGFLGISFLLGIALIVYKQYSEGHEDKRSYRILQEVGMSKSLKIKRTINSQIMIFF 589  
           LLF G + G F + A+ +YY+Q SEG D+ Y+ + ++GM+ K ++ I QI F  
   60   Sbjct: 515 LLFVGIIPTIFAILTAITYYQQSEGIKRDYKAMIKLGMNNTKIQDSIKVQINPVF 576

- Query: 590 FQRLVVAIVHFGVAIPMLKQMLVFGVINSTIYVVSGLTVAISIIYFIYRITSRTYY 649  
           P+ A+++ A+P+L ++ PG ++ + G + + Y+ I TS+ YY  
   65   Sbjct: 575 ILPIAPALNLIIFALPIYKIMITPGFNDAGLFLRAVGTCLIVYFVYFCHCTSKLYY 634

- Query: 650 HIER 654  
           +I +  
   70   Sbjct: 635 RLISK 639

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2115> which encodes the amino acid sequence <SEQ ID 2116>. Analysis of this protein sequence reveals the following:

-782-

Possible site: 35

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

|    |          |            |         |               |                        |
|----|----------|------------|---------|---------------|------------------------|
| 5  | INTEGRAL | Likelihood | =-13.59 | Transmembrane | 602 - 618 ( 592 - 630) |
|    | INTEGRAL | Likelihood | =-12.26 | Transmembrane | 59 - 75 ( 50 - 81)     |
|    | INTEGRAL | Likelihood | =-12.21 | Transmembrane | 235 - 251 ( 224 - 262) |
|    | INTEGRAL | Likelihood | =-9.82  | Transmembrane | 159 - 175 ( 146 - 177) |
|    | INTEGRAL | Likelihood | =-9.02  | Transmembrane | 201 - 217 ( 198 - 223) |
| 10 | INTEGRAL | Likelihood | =-8.97  | Transmembrane | 510 - 526 ( 507 - 540) |
|    | INTEGRAL | Likelihood | =-6.42  | Transmembrane | 569 - 585 ( 564 - 589) |
|    | INTEGRAL | Likelihood | =-5.95  | Transmembrane | 109 - 125 ( 102 - 138) |
|    | INTEGRAL | Likelihood | =-4.09  | Transmembrane | 294 - 310 ( 290 - 315) |
|    | INTEGRAL | Likelihood | =-1.86  | Transmembrane | 126 - 142 ( 126 - 142) |

----- Final Results -----

|    |                     |                                            |
|----|---------------------|--------------------------------------------|
| 15 | bacterial membrane  | --- Certainty=0.6434 (Affirmative) < succ> |
|    | bacterial outside   | --- Certainty=0.0000 (Not Clear) < succ>   |
|    | bacterial cytoplasm | --- Certainty=0.0000 (Not Clear) < succ>   |

The protein has homology with the following sequences in the databases:

|    |                                                                            |                                                                 |  |
|----|----------------------------------------------------------------------------|-----------------------------------------------------------------|--|
| 20 | >GP:BA03337 GB:AB035452 ABC transporter [Staphylococcus aureus]            |                                                                 |  |
|    | Identities = 141/657 (21%), Positives = 289/657 (43%), Gaps = 66/657 (10%) |                                                                 |  |
|    | Query: 5                                                                   | ITKENIKNFSLYRIYFLATIGLLSIFIAFLNPFISDKII--TEKIGDSQALVIANJSL- 61  |  |
|    |                                                                            | I M++N Y +Y L S+F + + + S + T + + + +I G+L                      |  |
| 25 | Subject: 6                                                                 | IVFYGLRQNLGHYAMV-----LPSLFSSIVLYPSFTTLOFTKGVNDDSAIIKKGALV 59    |  |
|    | Query: 62                                                                  | --IFLIIVFLVFLIYRNFFVKRQKLGVLALGFSKRELTGLLLENLIVLMSYLV 119       |  |
|    |                                                                            | IFL + +V+FL+Y N+ FVK+R++E + +G +++ +K+L LE +++ +++ +V           |  |
| 30 | Subject: 60                                                                | GSIFLPIIIVIFLHAYNHLVKRRTRFALPQLIGLTRQNLKMLALEQMVFLTGVGV 119     |  |
|    | Query: 120                                                                 | LLGPTLYFLAVLAITHLLNIMEVQWFTVNEIIESGLIVVFLINVTNGLIISKQS 179      |  |
|    |                                                                            | +L G L + ++ L+I++ + ++ ++I++ +++ + L + +S                       |  |
|    | Subject: 120                                                               | VLOGLAGAQLLSIVSKLMSLNLSHFEMALVITPMLIATYVLLPQALFKRRS 179         |  |
| 35 | Query: 180                                                                 | LIEFVNFPSKAS----KKIKIRKVAIIATALLSVILCLATYVFSSTNNMLSIGV 235      |  |
|    |                                                                            | ++ + S K + K ++ + I + I Y + +AT T L P                           |  |
|    | Subject: 180                                                               | ILSMKKDSIKTDATKAVTAEVISGLVIAIALGY--MATEMFTFKALINAWTSP- 236      |  |
| 40 | Query: 236                                                                 | SLLIIVLVLTGTVPTIRYGLAPVSVLLKNGKGLYRPLSNIIYKFNRYIATKMLTVP 295    |  |
|    |                                                                            | +I+ L V+G R ++ + EK++K + YR+ ET++                               |  |
|    | Subject: 237                                                               | -PIILFLTVVGAYLFFRSVSLIPKTLKKSNGRVSITDVPTSSIMRMKGNMSLTII 295     |  |
| 45 | Query: 296                                                                 | GGLLTIVTVSAGMVMMLAYSLNGIERLTPSAIEVNVSENGQVVTITILENDVSL-- 352    |  |
|    |                                                                            | + VTV+V + + + + + P+ R+V + T L Q++                              |  |
|    | Subject: 296                                                               | AIISAVTVTVLCFAALSKSNTDQTLTSMAPN--EFNVATQDAKQPEKLSQQQTSPSN 353   |  |
|    | Query: 353                                                                 | ----VDGGLRLNTIPEVTITDSGOTIPFDIINYSDYKLMKAQGRINSIRGSKSLPL 408    |  |
|    |                                                                            | + V ++ I +DGG+T N K G I +KSLP +                                 |  |
| 50 | Subject: 354                                                               | AYETITVDNVKDVITLNGSDGRTNSILSANN-----KVTGNALITNTKLPNI 405        |  |
|    | Query: 409                                                                 | INYYPTIEISLGRTPNLGNADVT--VKQSVNMFSESTSVITLV--VSDKLVAKLSSEF 464  |  |
|    |                                                                            | IN I L K + + T V Q V+ + S + V VS + Y +L +                       |  |
|    | Subject: 406                                                               | IN-----IHLAKDLVVKGTKNETFRVQEDKGRVYPLNLSFNSPVVHVSPEKYQLKLT-- 458 |  |
| 55 | Query: 465                                                                 | PEKEMTIRTPTNCTSI-----SSAFYNQFNMVSDVSSYKSHVKTIANIATYIFIT- 517    |  |
|    |                                                                            | + + TP G I+ ++A QF D + +Y + A IP+T                              |  |
|    | Subject: 459                                                               | ---QNNVHTFYGYDIKQTSQREKAQAIKQFG--DKVITYDEMKEVDRTNGILIPVTS 512   |  |
| 60 | Query: 518                                                                 | FLSLPIICTGSLIYPTSLIRIMNNKEYGLSKLGYSKRMHRIIRYTGTLPLPLPVI 577     |  |
|    |                                                                            | FL + P+ G I+Y + E + + L +G+ + + L + P+P+ I                      |  |
|    | Subject: 513                                                               | FLGLAFLVAAGCIYIKMDETEDELNFRILKRIQFHTHMLKGLIKITFNFGPLIJ 572      |  |
|    | Query: 578                                                                 | GIVNGMLLIYKYFLPMDTLVAGNIIMLSILLCLLPFLIYGTFFYVILTLVTSIIK 634     |  |
|    |                                                                            | I++ I + L GNI + +++ ++ +IY TP ++ +IK                            |  |
| 65 | Subject: 573                                                               | AIILHAVFAALAPMKIM-----GNISFMPVIVIVVYVTLIYITFALIAFVHNSKLIK 623   |  |

-783-

An alignment of the GAS and GBS proteins is shown below:

Identities = 145/678 (21%), Positives = 277/678 (40%), Gaps = 89/678 (13%)

```

5 Query: 13 NLKHSIDQYIPFLASLLYSITCSTL-----LIIMSAGVRDMGTATVFLGVIVSIF 67
 N+K + Y + LA++ L S+ + L I+ +G D G A + +I L +F
 Sbjct: 9 NIKKNFSLYRIYFLATIGLLSIFAPINFIIDKIITEKIG-DSGQALVIANGSLFLPLVIF 67

10 Query: 68 AVVMEHYSYINILMKQRSSSEGLYNILGWKQKQVAVASLRLFIYIFISIGSLPSAFA 127
 VV Y N +K+RS E G+ ILG +KR++ ++ +LE +I + L S
 Sbjct: 68 LVVFLIYFNFFVKKRSQELGVLAALIGFSKELTKLLLEINLVILV---LSVLVSLILG 123

15 Query: 128 KFTYILFVNIWNIHAINLSLMPFFIICIVFTIGFLTELVVPYIRHV-----RLSSPLS 181
 +Y + V I H LNL++ + FI I + + + V +I + + S +
 Sbjct: 124 FTYFLAVLAIT-ELNLITMEVQWFTITWKEIISLGIILVVFLINVTNGLIISKQSLTE 182

20 Query: 182 LFRKKQKQKEPEKGNLILALVALAIAYTMAL-----TSGKAPALAVIYRPFPAVL 235
 ++ EK+ K + AI+R+ A+ ++Y + L T ++ ++ ++L
 Sbjct: 183 FVNFSKASKKIKIRKVRATIAITALLLEYILCLATVFSSTRNMLSIGMVFVSELLITVL 242

25 Query: 236 VIAGTYLFYISPMYVGLRQRNKHYYIKSEHFVSTSQMIFRMQNAVGLASITLLAWA 295
 VA GT + + + L+ANK Y+ + + +R+ A +T+L +
 Sbjct: 243 VVLGTVFITYRIGLAFVSSLKENKGRVYRPLENIYKPKNYRI---ATNKNLTVLQGLL 299

30 Query: 296 LPTIATT---VSLYNTQNVVTLFPKFSVLSIGNSKGDAMNIFEKILKGLKSSKEAI 352
 VT++ V LY+ + N + L P ++ ++ G + + I
 Sbjct: 300 TVTVSVAGMVMILYAYSLNGIERLTPSAIEYNVESENQV-----NVTIT 344

35 Query: 353 TYNQTMISMPVQSSELNITSKVNKHVDITKTS---PHYLITQNDPRL-----GHQL 402
 N + + V + + + V IT +G + +I +D++ L + +
 Sbjct: 345 LENDQVSLVDVGL-----LRLNTIPEVITITDSQITIPYFDIINSYKELMKAQGRNTSI 399

40 Query: 403 PKLQMDQVAVFPQKGSRLKINLLGNKPNVQNLKAYVPETITYNPGLIFANNGKI 462
 K + + L K LGN +DV +K+ + +K
 Sbjct: 400 BGSKSLPLLINYPTEISLQKTFNLQNAVDT--VKQVSTNVSFSTSVTLVSDKLY 457

45 Query: 463 DNIRKAYLFPYTNINTFPKT-----FKAYLDLSQKINSISKNDIEVDGKYGVNIST 515
 + + + I TF T F + I+S SK ++ NI+T
 Sbjct: 458 AKLSRPFPEKMTIRTFNGTSIRSSEAFNQFSMPDVISYSKEHTVKT-----ANIT 512

50 Query: 516 KQSFLEKGYQMGGLLFTGFLIGSLFLGIALIVYKQYSEGHEDKRSYRILQEVMSKK 575
 +F FL I F++ I+Y+ E E+K Y L +G SKK
 Sbjct: 513 -----YIFITFL-SILFIICITGSLIYFSLIEIMENKEEYGLSKGLGYSKK 557

55 Query: 576 LVKRTINSQIMIFFPQPLVVAVHFGVAIPMLKQMLVFGVLNSTIVVVSGLTVLAISI 635
 ++ R + + I F P+ + ++ G+ + K L + ++ I + + L +L I
 Sbjct: 558 MIHRLRYSTGILFLIPFVIGIIVNGMMLIYYK-YLPMDTLVAGNIIMLSLLCLLFFLI 616

Query: 636 IYFIYRITSRTYTHIE 653
 IY Y +T R II+
50 Sbjct: 617 IYGFVFLVLRVTSIIK 634

```

A related GBS gene <SEQ ID 8639> and protein <SEQ ID 8640> were also identified. Analysis of this protein sequence reveals the following:

```

55 Lipop: Possible site: -1 Crend: 7
 McG: Discrim Score: -11.64
 GvH: Signal Score (-7.5): -3.52
 Possible site: 37
 >>> Seems to have no N-terminal signal sequence
ALOM program count: 11 value: -11.62 threshold: 0.0

60 INTEGRAL Likelihood = -11.62 Transmembrane 55 - 71 (49 - 75)
 INTEGRAL Likelihood = -10.30 Transmembrane 197 - 213 (192 - 218)
 INTEGRAL Likelihood = -9.13 Transmembrane 152 - 168 (141 - 172)
 INTEGRAL Likelihood = -8.70 Transmembrane 624 - 640 (619 - 645)
 INTEGRAL Likelihood = -8.44 Transmembrane 222 - 238 (219 - 250)
65 INTEGRAL Likelihood = -7.75 Transmembrane 283 - 299 (280 - 307)

```

-784-

|   |                           |                    |               |                        |
|---|---------------------------|--------------------|---------------|------------------------|
|   | INTEGRAL                  | Likelihood = -7.70 | Transmembrane | 533 - 549 ( 526 - 552) |
|   | INTEGRAL                  | Likelihood = -6.95 | Transmembrane | 108 - 124 ( 99 - 140)  |
|   | INTEGRAL                  | Likelihood = -4.88 | Transmembrane | 585 - 601 ( 581 - 610) |
|   | INTEGRAL                  | Likelihood = -3.82 | Transmembrane | 25 - 41 ( 21 - 47)     |
| 5 | INTEGRAL                  | Likelihood = -0.48 | Transmembrane | 602 - 618 ( 602 - 618) |
|   | PERIPHERAL                | Likelihood = 1.16  | 129           |                        |
|   | modified ALOM score: 2.82 |                    |               |                        |

\*\*\* Reasoning Step: 3

----- Final Results -----

```
bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

ORF02245(310 - 2262 of 2562)  
 GP[380256]sb[AA969695.1]AF267498\_5[AF267498(1 - 639 of 640) permease OrfY {Streptococcus  
 mutans}  
 %Match = 10.2  
 %Identity = 24.0 %Similarity = 49.8  
 Matches = 147 Mismatches = 297 Conservative Sub.s = 158

25 123 153 183 213 243 273 303 333  
QKTC\*IVLKLTLWMDKLF\*W\*PIQQMLLVMPNAPFYLKMDVFFTFNFIVVIRIIANSIKIFL\*QCLPY\*GVNNPFYLIKAW  
|||:::  
MFLPKISF

30            363            393            423            453            474            504            534            564  
HNKHSIDQYIPFLLASLLYSCTSTLLILMSAVGRDMGTAT---VLFGVILSIFAVWMEHYSYNLMQRSSEFG  
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
HNLIVNKSLYLFPYAIMTISGPNLYNLFINPSYNIPTARILLIDILIGFLISLMLILGYRANFYSDERNNGM  
               20                30                40                50                60                70                80

[illegible]

834            864            894            924            954            1011            1041  
PVIRHVLHSPLSLFRKQOQEKEPKENLILAILVAIAIATYTTALTSGKAPALAVITY-RFFFAVLLVIAGTYLFYISF  
     | :         | :         | :         | :         | |||         | :         | :         | :         | :  
45 SEYRLKKQSIIVIPNISKARNRNPRTSVPVGGLGFLFALLMGYHFALTS---PNVTTSFSRFTYAACLVLGIIGTCPTSSG  
     180          180          180          180          180          240          240

[illegible]

1311            1341            1371            1401            1431            1461  
 SKGDAGNIFEEKILKKLGKSSKEATTYNQIMISMPVSSSELMITSNKVGVINDITTKGFM-----  
 ||| : : : : | : : : || :|| : : : :  
 -----TVLSTTIIIEPNMADIATIN-----HVTLTNRNQLKVSQSVYGNIGKSLSVDPEN-----  
                        320                  330                  340                  350                  360

[illegible]

65            1698          1728          1758          1788          1818          1842          1872          1902  
NNKQI-DNIRKAYLPYTIKININTFPRTFKAYILDINSQEINSISKNDIIEVDG--KYVGNISTKQSFLKEGYCMFGGLLFTIG  
|::|| | : | : : || : | :: |: : : : |||||

-785-

```

DNREIIQTILKRELTWG-----TMAGY-HVKGKKNQKDFYDELSTNFRQFSANVVSIRQVKSFNALFGGLLVG
 460 470 480 490 500 510
5 1932 1962 1992 2022 2052 2082 2112 2142
 FLGIGISPLLGIALIYVYEQYSEGHEDKRSYELLQVQMSKKLVKRTINSQIMIFFPQQLAVAVITHFGVALDMLKQMLLVF
 :: | : | : | | | : | : : | | : : | | | : | : : : | : | : : |
 IIPGTIFALITAITIYQQLSEGISRDSDYKAMIKLGMINKTIQDSIKVQLNFVFLIPALPALLGLPALPILYKIMTFP
 530 540 550 560 570 580 590
10 2172 2202 2232 2262 2292 2322 2352 2382
 GVLNSTIVYVVGSLVLAISIIYFIITRITSRTIYHIER*KGLVILPILH**KPID*KICYTK*KKESIVYPRGTVT
 | : : | : | : | : | | : | : |
 GFNDAGLFLRAVGTCLIVVLFYWFICHCTSKLYRLLSKK
 610 620 630 640
15

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 687

A DNA sequence (GBSx0729) was identified in *S. agalactiae* <SEQ ID 2117> which encodes the amino acid sequence <SEQ ID 2118>. This protein is predicted to be ABC transporter OrfX. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5121 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF99694 GB:AF267498 ABC transporter OrfX [Streptococcus mutans]
Identities = 118/242 (48%), Positives = 175/242 (71%), Gaps = 1/242 (0%)

Query: 5 INHLEKVFRTPSKETRALQVDVFKVQGEFTIAGCESGSGRTTLNLAILEKPTNGQ 64
 ++HL+KV++T+ AL+D+ F V++GEFTIAGCESGSGK+TLNLA ++ P++G
Sbjct: 6 VSELKKVYKQEGLTN-BALKDITFSVQEGFTIAGCESGSGKSTLLNLAICHNDYPSGGH 64

Query: 65 VILNGEDITKIEAKLASFELONLGFVPQDFNLLDLSVRDNIYLFIVLDRKRYKEMCHR 124
 + I N + K+K+ + A FR ++GF+PQ+FNLL+ + +DN+ +P+++ + + R
Sbjct: 65 IIFNNYQLEKVKDEAAVFRSHIGFIFQPNFLLNIFNNKDNLLIPVIISSGSKVNSYEER 124

Query: 125 LSELSSHLRIDDLDRPFELSGGQKQVARSALITNPQILLADEPTALDYNSEDLL 184
 L +L++ + I+ IL K P+ELSGGQ+QR+AIAR+LI NP ++LADEPT LD + S+ +L
Sbjct: 125 LRDLAAVGVIESILSKYPYELSGGQQRATARALINNPDLILADEPTGQLDSTSRQL 184

Query: 185 NLFETINLDGQITLMTVTHSANAASHAKRVLPFKDGRIFHQLYRGNKNNSFPNKDISLIMS 244
 NL IN +TILMTVTHS AAs+A RVLPIKDG IP+QL RG K+ P I + +
Sbjct: 185 NLLSNINAKRKTILMTVTHSPKASVYANRVLPFKDGVIPNQLVRGCKSRBGLDQIIMQA 244

Query: 245 AI 246
 ++
Sbjct: 245 SL 246

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2119> which encodes the amino acid sequence <SEQ ID 2120>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

-786-

bacterial cytoplasm --- Certainty=0.2131(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 91/222 (40%), Positives = 142/222 (62%), Gaps = 2/222 (0%)

Query: 2 LLEINHLEKVFTRFRFSKESTRALQDVDFKVBQGEFTIAMBSGGGKTLINILATLEKPT 61  
 LL + + K + SE L+ +D +V +G+P+AIMG SGGGK+L+NI+ L+K+  
 10 Sbjct: 1 LNLKIDIRSYH--LQTFEFAILKIGIDLEVNEGDFIAIMGPGSGGKSTLMNIGCLDKEG 58

Query: 62 NQQVILNGEDITIKIEAKLASPRLLKNGLFVQDFNLDTLSVRNDIPLVLDRKRYKEM 121  
 +G + G D+ + + +LA R + +GFVQ+NL+ L+ N+ LPL KE  
 15 Sbjct: 59 GGSVAIEGRDVSLSLSDNLAIDLNRQKIGFVQNNMLPKLTACQNVLPPTYMNVFKKER 118

Query: 122 DFRLSSELSSHLRIDLLDKRPFELSGGQKQKVAIARSLITNPQILLADEPTAALDYRNS 181  
 R E+ + + + + +P ELSSGGQKQKVAIAR+L+TNP +L DEPT ALD + S  
 Sbjct: 119 RKRALMLKLVGLSEKSEFKPMELSGGQKQKVAIARALVNPSFILSDEPTGALDTKTSV 178

Query: 182 DLLNLPETINLDGGQTLIMVTHSANAASHARVLFKDGRIHF 223  
 +++LF+ N +G+T+++TH A+ K+ + +DG I H  
 20 Sbjct: 179 QIMDLFKQFNDNGKTIITITHEPEVAALCKTVILRDGNIEH 220

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 25 vaccines or diagnostics.

#### Example 688

A DNA sequence (GBSx0730) was identified in *S. agalactiae* <SEQ ID 2121> which encodes the amino acid sequence <SEQ ID 2122>. This protein is predicted to be nisin-resistance protein. Analysis of this protein sequence reveals the following:

30 Possible site: 18  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -13.16 Transmembrane 8 - 24 ( 1 - 31)  
 ----- Final Results -----  
 35 bacterial membrane --- Certainty=0.6265(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:AA08491 GB:U25181 nisin-resistance protein [Lactococcus lactis]  
 Identities = 108/318 (33%), Positives = 190/318 (58%), Gaps = 8/318 (2%)

Query: 3 RKIVLFPVPMILIVLGLGVVVHYGSAINIYLLPSSERYGRVILDRVQGRGLYSQGRQ 62  
 ++I+L V + LGI +++G NIYL+PES +Y RV L + ++ GL++ ++  
 45 Sbjct: 5 KRILLGLVAVCALFLGI----TYFWGYKFNILYVPPSPQKYVRVALKIMDELGLFTDSKE 60

Query: 63 WQIIQRSEKKLKTSSKYOSSRNIVQEAIVRGYGGGHSQILSKETVRRDITLDSRYPEYRRL 122  
 W ++++ ++ +K+Y E+ +Q+A++ GKGIS I +E + + ++ +  
 50 Sbjct: 61 WVEETKKKTIETSSNAKNYASTIPLQKAIKAVAGGKHSFIEHEEDIKSRSTKYIKPKAEI 120

Query: 123 NEDILLITTPISIKLDRKSISHYSGKLQNTLMKESYKGLILDLSNNTGGMIMPIGGVAS 182  
 + L++TIP + D ++ S Y+ L+ + +Y G+I+DL N GG++ PM+ G+  
 Sbjct: 121 EGNVTLITLITPEPTGNDGQA-SDYANPLRSSPHKNVYGVVDLRGNGGDLSPWVLGLSP 179

Query: 183 ILPNDTLPHYTDKYGNKKTITMKNIPILKALKISKRTINTKHV---PIAIIITHKTSASAS 239  
 +LP+ TLF Y DK + K + +N + + S K + K + PIA++ + T SS E  
 55 Sbjct: 180 LLEFDGTLFTYVDKSHSKPVELQNGHNSGSSSTKVSNDKKIKKAPLAVLINNTGSSGE 239

Query: 240 MTFLEFKLEPNVKSFGQATAGYITVNETPMIYDGARIALTGIIVSDRQGYKYENTPILPD 299  
 +T L FKG+PNVK G +AGYT+ N+T LYDG+ L+T+ V DR Y+N PI PD  
 60



-787-

Subject: 240 LTALCFKGIPNVKFLGSDSAGYTSANQTVYLYDGSTLQITSAFVKDRINNIYKNFPISPD 299

Query: 300 QVTSLPLQESQSWLKSRI 317

$$T_t + W_t K S_t I$$

5           Spict: 300 IOTNNAKSSAIEWIKSOI 317

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8641> and protein <SEQ ID 8642> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1    Crend: 3

McG: Discrim Score: 12.71

GvH: Signal Score (-7.5): -5.64

Possible site: 18

>>> Seems to have an uncleavable N-term signal seq

```
ALOM program count: 1 value: -13.16 threshold: 0.0
```

INTEGRAL Likelihood = -13.16 Transmembrane 8 - 24 ( 1 - 31)

|            |                   |     |
|------------|-------------------|-----|
| PERIPHERAL | Likelihood = 4.03 | 174 |
|------------|-------------------|-----|

modified ALOM score: 3.13

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.6265(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

34.7/62.5% over 311aa

Lactococcus lactis

GP|805128| nisin-resistance protein Insert characterized

ORF01108 (343 - 1254 of 1560)

GF|805128|gb|AAB08491.1||U25181(7 - 318 of 318) nisin-resistance protein {Lactococcus lactis}

%Match = 19.4

%Identity = 34.6    %Similarity = 62.4

Matches = 106 Mismatches = 112 Conservative Sub.s = 85

[illegible][illegible][illegible][illegible]

1164 1194 1224 1254 1284 1314 1344 1374  
TGIVSDROGYKYENTPILFDQVTSLPLOESQSWLKSRIMN\*GIINKGELYVIRNQLRKSFSYTFKRRDKGSTRRRF

-788-

```

|: | || | | | | | | | | | | | | |
TSAFVDRITNINYSNFFISPDICITNNAKSGATEWIKSQIK

```

SEQ ID 2122 (GBS38) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 12; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 689

A DNA sequence (GBSx0731) was identified in *S.agalactiae* <SEQ ID 2123> which encodes the amino acid sequence <SEQ ID 2124>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

```

```

----- Final Results -----

```

```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2125> which encodes the amino acid sequence <SEQ ID 2126>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.1369 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 31/49 (63%), Positives = 43/49 (87%)

Query: 6 KKLTKSLPGIKLISIIPTTELIGKADINSRPIIEKELDRRHEKKIDL 54
 K++ K+LG +GKL+SI+PDTTE+IGK IDNSRPIIEK ++++HEK+ L
Sbjct: 3 KIRKALGVGKLMISIVPTTELIGKTIIDNSRPIIEKRMCKHEKEMQL 51

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 690

A DNA sequence (GBSx0732) was identified in *S.agalactiae* <SEQ ID 2127> which encodes the amino acid sequence <SEQ ID 2128>. Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.3644 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 2126.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 5 Example 691

A DNA sequence (GBSx0733) was identified in *S. agalactiae* <SEQ ID 2129> which encodes the amino acid sequence <SEQ ID 2130>. This protein is predicted to be 28 kd outer membrane protein precursor (yaeC). Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> May be a lipoprotein

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAE59827 GB:J012388 hypothetical protein [Lactococcus lactis]
Identities = 123/290 (42%), Positives = 178/290 (60%), Gaps = 18/290 (6%)

20 Query: 1 MKIKKLLGLTTTVVISALILGAC-----GQSINEDAKVVRVGTMVSKTEKARMDKIEE 54
 +K +++L +T +++ +I+G G +K+V++G M K E M +++
Sbjct: 3 VGRHIL-ITIIILVFIIVGIGIFAFSHSGNKVSSKIVKIGLMPGGKQEDVTKQVQK 61

25 Query: 55 LVKKK-GVKLKFTPTDYTOPKFALESDEIDINAPCHYMLNWNKANKITMLVSVAETIF 113
 K + G+ LKF PTD +PKAL +E+D+NAQFY YL +NKAH N+V8+ +T
Sbjct: 62 NAKDQPGITLKFNVPNTDGDENKALVNHVEDLNAFQEHAYILKSNKANKNIVSIGDTII 121

30 Query: 114 TSFRLYSGTNGKRGKYQTVSEIPNKAITTPMDAVNERRSLYLLQSAGLLKLKVSODALA 173
 T LVS KY+ V EIP+K+TI IEND NERR+LY+L+AGL+KL S LA
Sbjct: 122 TPIHLYST-----KYKVDRIPOKSTIATPNDITNERRALYVKNAGLIKIDTSRGVLA 175

35 Query: 174 TMSDVVSNPKSLDLKEVDAACTARSLOSTDAVINDFFTEAGINPKSAIPIEPKSKNAK 233
 2+ D+ NRESL +KE+DA+QT R+LDS AAVIN +F A + K +I+ EP +++
Sbjct: 176 TVKDIRENPKSLIIEKIDASQTPRALDSVAAVINDYINFAISAKNSKESYIQRFPLNDS 235

Query: 234 QWYNLLVAQKQMDCKKAKAIKEVVKAYHTDAVKVIEKT-SCGLDQPW 282
 QW N + A Q K KEVVKAY + +I+K G + P W
40 Sbjct: 236 QWYNFIAAN---QSDKNNKVVKEVVKAYEQKNIAIDITKKEYPDGGELPW 282

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2131> which encodes the amino acid sequence <SEQ ID 2132>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1766 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 145/264 (54%), Positives = 203/264 (75%), Gaps = 2/264 (0%)

55 Query: 20 LGACGQSINEDAKVVRVGTMVSKTEKARMDKIEELVKRGVVKLFTEFTDYTOPKFALES 79
 L C + K +D + +G M K++++ARWDK+EEL+KK + LK+ EFTDY+QPKA+
Sbjct: 1 LVACSE-KQDDKNITLIGVMTKTESQARWDKVEELKDKNITLKYKEFTDYSPQKAVA 59

```

-790-

Query: 80 EDEIDINAFQHYNVLANWANKANKTNLVSVAETYPFSRLYSGT-KKGGKGYQTVEIPIPK 138  
 + E+DINAFQHYN+LANWANK NK +LV++A+TY + L+SGT ++GK KY++V++PN  
 Sbjct: 60 NGEVDINAFQHYNVLANWANKENKEHLVAIADTYIETINLPGTSGDGGKAKYKSVADLPWG 119

5 Query: 139 ATTTTPNDVAVNESRLYLQSGAGLLKLVSGDALATMSEDVVENPKSLDLKRVDAQTARS 196  
 I +PNDA NESR+LY+LQSGAG+KL VSGD LVT++++ N K LD+KE+DA+QTAR+  
 Sbjct: 120 TQIAPVNDATNESRALYVLQSGAGLIKLVSGDQLATIANISINKKLDYKRLDASQTARA 179

10 Query: 199 LDSTDAVINNDPVTGAINPKSAIFIEPKSKMAQWYNLLVAQRGWQKSKAKAIKEVV 258  
 L S DAIV+NN + A I+ K+++F E N+KQW N++ QK W+ KA AIK+++  
 Sbjct: 180 LVSADAIVVNSYAVPAKIDYKTSLFKEKADNRSQWINTIAGQDWBKSEKADAIKKLI 239

Query: 259 KAVHTDAVKKVIEKTSQGLDQPVW 282  
 KAY TD VKKV+EKTS G+D VW  
 15 Sbjct: 240 KAYCTDEVKKVVERTSKGILDSVW 263

SEQ ID 2130 (GBS96) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 3; MW 57.2kDa).

- 20 The GBS96-GST fusion product was purified (Figure 195, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 290), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 692

A DNA sequence (GBSx0734) was identified in *S.agalactiae* <SEQ ID 2133> which encodes the amino acid sequence <SEQ ID 2134>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.5103(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 9807> which encodes amino acid sequence <SEQ ID 9808> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 693

A DNA sequence (GBSx0735) was identified in *S.agalactiae* <SEQ ID 2135> which encodes the amino acid sequence <SEQ ID 2136>. This protein is predicted to be glucose-inhibited division protein (gid).

- 45 Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> Seems to have no N-terminal signal sequence

-791-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0656 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:CB13486 GB:Z99112 glucose-inhibited division protein [Bacillus subtilis]  
 Identities = 289/439 (65%), Positives = 352/439 (79%), Gaps = 10/439 (2%)

10 Query: 1 MSQSYNVIGAGLAGSEAAVQIAKRGIPVKLYEMRGVSKTPQHKTTNFAELVCSNSFRGD 60  
 M+Q +NVIGAGLAGSEAA+QAKRGI VKLYEMR VK TP H T FAELVCSNS R +  
 Sbjct: 1 MNQQTVMVIGAGLAGSEAAWQIAKRGICQVKLYEMRPFVKTPAHHTDKFAELVCSNSLRSN 60

15 Query: 61 SLTNAVGLLKEEMRPLDSIIMRNGEAKRVPAAGGNAVDRRGYSEAVTEIHKHPLIEVIR 120  
 +L NAVG+LKEEMR LDS I+ + VPAGG+AVDR ++ +VT + HD + VI  
 Sbjct: 61 TLNAVGLVKEEMRALDSIIAANDECSVPAGGALAVDRHEAASVTVNRVKNHFWTVIN 120

20 Query: 121 DEITDIPGDAITVIATGFLTSDSLAAKIHENCGSDGPFYDDAAPIVDKNTIDINKVYLK 180  
 +E+T+IP + T+IATGFLTS+SL+AA+ EL G D YPDAAAPIVK++D++KVYLK  
 Sbjct: 121 EEVTEIP-EGPTTIIATGFLTSESLAQLKELTGEDYLYPYDAAPIVEKDSLMDKVKYLK 179

25 Query: 181 SRVDKGEAAYLNCMP+KEEPMFAPHEALTAREAPINSPEKEYPFEGCMPIEVMAKRGIKT 240  
 SRVDKGEAAYLNCMT+EEF PHEALT+AE PL FEKE +FEGCMPIEVMAKRG K T  
 Sbjct: 180 SRVDKGEAAYLNCMT+EEEPDRPHEALTSAETVPLKEFEKEIFPEGCMPIEVMAKRGKKT 239

30 Query: 241 MLVGPMPKPVGLIEPYEDYKGRDGEPTKPYAVVQLRQDNAGSLYNVGPQTHLWGEQNR 300  
 ML+GMPKPVGLIE+P K PYAVVQLRQD+AA+LYNIVGPQTHLWAG+CK  
 Sbjct: 240 MLPGMPKPVGLIEHPVTGK-----RFPYAVVQLRQDAAGTLYNIVGPQTHLWGEQKE 291

35 Query: 301 VFQMIPLGENAEFVRVGVNHRNYSYMDSEPNLLNTQFATRNPNLFPAGQMTGVGEYVESAA 360  
 V++IPLGEN E VRYGVNHRN++SP+LL T+ + +LFPAGQMTGVGEYVESAA  
 Sbjct: 292 VLKLIPLGENVEIVRYGVNHRNT+FINSPSLKITYQFKNRSDLPFAGQMTGVGEYVESAA 351

35 Query: 361 SGLVAGINAVRRFNGSESVVFPQTATGALPHYITHTDSKHPQPMNVNFGIIEKLEGPRI 420  
 SGLVAGINA + GE V+FPQ TAIG++ HYIT T+ K+FPQMN NFG++KEL +I  
 Sbjct: 352 SGLVAGINAAKLVGEELVIFPQETAIGSMHYIITTNQKNFQPMNANFOLLKELF-VKI 410

40 Query: 421 RDKKERYEAIAITRAKIDLE 439  
 ++KKER S A RA++ ++  
 Sbjct: 411 ENKKERNEQYANRAETIQ 429

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2137> which encodes the amino acid sequence <SEQ ID 2138>. Analysis of this protein sequence reveals the following:

Possible site: 30

45 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -8.44 Transmembrane 12 - 28 ( 9 - 32)

----- Final Results -----

bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55

RGD motif: 111-113

The protein has homology with the following sequences in the databases:

>GP:CB13486 GB:Z99112 glucose-inhibited division protein [Bacillus subtilis]  
 Identities = 292/435 (67%), Positives = 350/435 (80%), Gaps = 10/435 (2%)

60 Query: 59 INVIGAGLAGSEAAVQIAKRGIPVKLYEMRGVSKTPQHKTTNFAELVCSNSFRGDSLINA 118  
 +NVIGAGLAGSEAA+QAKRGI VKLYEMR VK TP H T FAELVCSNS R ++L NA  
 Sbjct: 6 VNVIGAGLAGSEAAWQIAKRGICQVKLYEMRPFVKTPAHHTDKFAELVCSNSLRSLTANA 65

Query: 119 VGLLKEEMRPLDSIIMRNGEAKRVPAAGGNAVDRRGYSEAVTEIHKHPLIEVIRGEITE 178

-792-

VG-LKEEVR IDS I+ + VPAGGA+AVDR +A SVT ++NHP +VI E+TE  
 Sbjct: 66 VGV-LKEEVRALDSIAIIAADECSVPAGGALAVDRHEFAASVTNKVKHPRVTVINEEVT 125

5 Query: 179 IPDADITVIATGPLTSDLAELKIHALNGDGFYFYDAAAPIDKSTIDMSKVYLKERYDK 238  
 IP+ T+IATGELTSS++L+ ++ L G D YFYDAAAP++K ++DM KVYLKERYDK  
 Sbjct: 126 IPGEF-TIIATGPLTSELSAQLELTGCEYLYFYDAAAPTEKDSLDNMKVYLKERYDK 184

10 Query: 239 GEAYALNCMTKEEFMAFHEALT+TAEAPLANAPEKEKYFEGCMPIEVMKRGIKTMLYGP 296  
 GEAYALNCMT+EEF FHEALT+AE FL FEKE +FEGCMPIEVMKRGIKTML+GP  
 Sbjct: 185 GEAYALNCMTTEEFDRFHEALTSAETVFLPEKEEIFEGCMPIEVMKRGIKTMLPGP 244

15 Query: 299 MKPVGLEVPDDYTGPRDGEFKTPYAVVQLRQDNAGSLNINVGPCQTHLKGQKRVFPQ 358  
 MKPVGLE+P TG R PYAVVQLRQD+AGS+LYNINVGPCQTHLKG+OK V ++I  
 Sbjct: 245 MKPVGLEHP--VTGKR-----PYAVVQLRQDAGSTLYNINVGPCQTHLKGQKQEVKLI 296

20 Query: 359 PGLNLAEFVRYGVMMHNSYMDSPNLLTETPQSRSNPMLFPAQMTGVGYSVSAAGSLVA 418  
 PGLN E VRYGVMMHNS+++SP+LL TQ ++ +LFPAGQMTGV+GYVSAAGSLVA  
 Sbjct: 297 PGLNVEIVRYGVMMHNTFINSPELLKPTTYQKRNRSDFLPAQMTGVGYSVSAAGSLVA 356

25 Query: 419 GINAAALFKREBALIFPQTTAIGSLPHYVTHADSKHFQPMNVNPGIIELEGPRIIDK 478  
 GINAA+L EE +IFPQ TAIGS+ HY+T + K+QPMN NPG++KEL ++I+KKE  
 Sbjct: 357 GINAAKLVLEELVIFPQTTAIGSMHYTTTTNQNFQPMNVNPGIIELEGPRIIDK 415

25 Query: 479 RYEALASRALDLDT 493  
 R E A+RA+ + T  
 Sbjct: 416 RNEQYANRAIETIOT 430

An alignment of the GAS and GBS proteins is shown below:

Identities = 395/439 (89%), Positives = 417/439 (94%)

30 Query: 4 SYINVIAGLAGSEAAQIAKRGIPVKLYEMRGVKSFTQKHTDNFAELVCSNSFRGDSLT 63  
 +YINVIAGLAGSEAAQIAKRGIPVKLYEMRGVKSFTQKHT DNFAELVCSNSFRGDSLT  
 Sbjct: 57 TYINVIAGLAGSEAAQIAGRGIPVKLYEMRGVKSFTQKHTDNFAELVCSNSFRGDSLT 116

35 Query: 64 NAVGLLKEEMRLDSIIMRKGZHRVPAAGAMVDREGYSEAVTEIHGHLIEVIRDEI 123  
 NAVGLLKEEMRLDSIIMRKGZ+RVPAAGAMVDREGY+E+VT E+ HPLIEVIR E+  
 Sbjct: 117 NAVGLLKEEMRLDSIIMRKGZHRVPAAGAMVDREGYAESVTAELNHPHLEIVIRGEI 176

40 Query: 124 TDIPGDAITVIATGPLTSDSLAAKIHALNGDGFYFYDAAAPIDKSTIDMSKVYLKERY 183  
 T+IP DAITVIATGPLTSD+LA KIH ALNGDGFYFYDAAAPIDK+TID+KVYLKERY  
 Sbjct: 177 TEIPDADITVIATGPLTSDSLAAKIHALNGDGFYFYDAAAPIDKSTIDMSKVYLKERY 236

45 Query: 184 DKGEAYALNCMTKEEFMAFHEALT+TAEAPLANAPEKEKYFEGCMPIEVMKRGIKTMLY 243  
 DKGEAYALNCMTKEEFMAFHEALT+TAEAPLAN+PEKEKYFEGCMPIEVMKRGIKTMLY  
 Sbjct: 237 DKGEAYALNCMTKEEFMAFHEALT+TAEAPLANAPEKEKYFEGCMPIEVMKRGIKTMLY 296

50 Query: 244 GPMKPVGLEVPDDYTGPRDGEFKTPYAVVQLRQDNAGSLNINVGPCQTHLKGQKRVFPQ 303  
 GPMKPVGLEVP+DY GPRDGEFKTPYAVVQLRQDNAGSLNINVGPCQTHLKGQKRVFPQ  
 Sbjct: 297 GPMKPVGLEVPDDYTGPRDGEFKTPYAVVQLRQDNAGSLNINVGPCQTHLKGQKRVFPQ 356

55 Query: 304 MIPGLNLAEFVRYGVMMHNSYMDSPNLLTETPQSRSNPMLFPAQMTGVGYSVSAAGSL 363  
 MIPGLNLAEFVRYGVMMHNSYMDSPNLL+TF +R NPNLFPAGQMTGVGYSVSAAGSL  
 Sbjct: 357 MIPGLNLAEFVRYGVMMHNSYMDSPNLLTETPQSRSNPMLFPAQMTGVGYSVSAAGSL 416

60 Query: 364 VAGINAVRRNCESEVVPQTTAIGALPHYIHTDSKHFQPMNVNPGIIELEGPRIIDK 423  
 VAGINA R F E ++FPQTTAIG+LPHY+TH DSKHFQPMNVNPGIIELEGPRIIDK  
 Sbjct: 417 VAGINAAALFKREBALIFPQTTAIGSLPHYVTHADSKHFQPMNVNPGIIELEGPRIIDK 476

Query: 424 KERYEALATRALDKLEKFL 442  
 KERYEAL+RAL DL+ L  
 Sbjct: 477 KERYEALASRALDLDTCL 495

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 694**

A DNA sequence (GBSx0736) was identified in *S.agalactiae* <SEQ ID 2139> which encodes the amino acid sequence <SEQ ID 2140>. This protein is predicted to be transcriptional regulator (GntRfamily). Analysis of this protein sequence reveals the following:

Possible site: 13  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.5103(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04138 GB:AP001508 transcriptional regulator (GntR family)  
[Bacillus halodurans]  
Identities = 83/229 (36%), Positives = 133/229 (57%), Gaps = 1/229 (0%)

Query: 2 LPAYIKHDAIKKSIDKWTGKIGQRLPSERLDADYSVSRMTLRQSITLLVSEGLERRV 61  
LP Y +I + IK++I+ G K G L SER+ A+ Y VSRMT+EQ+I LV +G + +  
Sbjct: 6 LPIYQIQIEKQQLBSGVLRKPGDMLKSEREYAEYDVSRMTVRQAINNLVNGQYIYKKK 67

Query: 62 GSSTYVASHRVQKRMGGTTSPTIEVNSQGRKPSKLLSPQRKLANSTEIQLNLSQSDYV 121  
GSSTYV +++ + G TSPTE + +G +PSG+L+ F+ A ++LNL ++ V  
Sbjct: 68 GSSTYVQKKKIEQALNGLTSPIDMRKKGMEPSSRLKPELI PATAKIAKELNLKENTFV 127

Query: 122 VMREKRVYADKVPVLVYVVASIPENLIGFQSGEVTEHFFKTLTEN-GYEIGKSQQTIYAR 180  
+R+RY D VP+ E +P NL+KG + + + + E I + Q I A  
Sbjct: 126 TEIKIRRYGDGVFAIEKKNLLEPANLVKGLNKEIINQSLYQYIEELNLRADALQVIEAS 187

Query: 181 NASERVASHLEVNAGHAILALTQVSYPFDGKPEPVYHQQYVDRGFEEFL 229  
AS+ A LE+ G IL + + ++ DG E V Y DR++P +  
Sbjct: 188 TASKTEADLLEIQKGSPIILLERKTFADGTVLELVKSAYRADRYKPMI 236

There is also homology to SEQ ID 1256.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 695**

A DNA sequence (GBSx0737) was identified in *S.agalactiae* <SEQ ID 2141> which encodes the amino acid sequence <SEQ ID 2142>. This protein is predicted to be GMP synthase (guaA). Analysis of this protein sequence reveals the following:

Possible site: 46  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -0.96 Transmembrane 228 - 244 ( 228 - 245)

----- Final Results -----  
bacterial membrane --- Certainty=0.1383(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD15805 GB:AP056326 GMP synthase [Lactococcus lactis]  
Identities = 416/511 (81%), Positives = 467/511 (90%), Gaps = 3/511 (0%)

Query: 10 IQKIIVLVYGSQYNQLIARRIRIEGVFSLELKHKITADEIRDINPIGIVLSGGPNSVYD 69  
++KIIVLVYGSQYNQLIARRIRIEGVFSLELKHKITADEIRDINPIGIVLSGGPNSVY +  
Sbjct: 6 LEKIIVLVYGSQYNQLIARRIRIEGVFSLELKHKITADEIRDINPIGIVLSGGPNSVYD 65

Query: 70 GAFGIDEEIFELGIPILGICYQMQLITHKLGKVLPAEGAGHRYQGSALRLRESALFA 129  
 G+F ID EIPELG+P+LGICYQMQL+++KLG V AG+ REYG + L+L +SALFA  
 Sbjct: 66 GSFIDPEIFELGLFVLGICYQMQLMSYKLGWVEAGE---REYGVAPQLTEKSALFA 122

5 Query: 130 GTFQEQVLMSHGDAVTEIPESPHLVGSDNCPFAIENTEKNLYGIQPHPEVRHSVGN 189  
 GTP+ Q VLMSHGDT IPESPH+VG S + PFAA+ENTE+ YGIQPHPEVRHSV+G  
 Sbjct: 123 GTFPEQDVLMSHGDRVTALPEGPHVVCTSPNSPFAAVENTERNLYGIQPHPEVRHSVGT 182

10 Query: 150 DILKIFPAVNICARGDWSMDNFIDMETAKIRETVGDRKVLGLGGVDSSVGVLLQRAI 249  
 ++L+NFA++ICGA+G+WSM+NFIDM+I IRE VGD+KVLGLGGVDSSVGVLLQRAI  
 Sbjct: 183 EMLRNFALNICAGKWSMNFIDMQIKDIRKVGDKVLLGLGGVDSSVGVLLQRAI 242

15 Query: 250 GDQLTCTFVDHGLLRHNEGDQWMDLGGKPLHAIIRVDASKRFLDLLSGVEDPERKKLI 309  
 GDQLT IFVDHG LRK E DQWM+ LGKPLHAI+VDA KRP+D L G+ DEF +KKII  
 Sbjct: 243 GDQLTSIFVDHGLRKGADQWMTLGGKPLHAIKVDQKRPMDLVGLSDPETQRKII 302

20 Query: 310 GREFVTVFDEASKLGVDFLAQGLTYTDIIBSGTETAQTIKSHHNVGGLPEMDQFELIE 369  
 GMEFVTVFDEEA+KL+GVDFLAQGLTYTD+IBSGT+TACTIKSHHNVGGLPEMDQF+LIE  
 Sbjct: 303 GREFVTVFDEANKLGVDFLAQGLTYTDVIBSGTDTAQTIKSHHNVGGLPEMDQFOLIE 362

25 Query: 370 PLNTLPKDEVRALGTALGMDEIVNRQPPFGGLAIRVMEITEEKLTVRESSDAIRRE 429  
 PLNTLPKDEVRALGT LGMDE+VNRQPPFGGLAIRV+G+TEEKLTVRESSDAIRRE  
 Sbjct: 363 PLNTLPKDEVRALGTQLGMDEIVNRQPPFGGLAIRVLGDLTEEKLTVRESSDAIRRE 422

30 Query: 430 IAKGALDRDVMQYPTVNTVKSQVSGVMDGRTDYDTIAIRAITSIDGMTADFAQLPMDVILK 489  
 IA +GL+RDVMQYPTVNT V+SVGVMGD RTYDVT+AIRAITSIDGMTADFAQLPMD+L+  
 Sbjct: 423 IAASGLERDVMQYPTVNTVKSQVSGVMDGRTDYDTIAIRAITSIDGMTADFAQLPMDVILQ 482

30 Query: 490 KISTRIVNEVDHVRIVVDITSKPPATVEMQ 520  
 KIS RIWNEVDHVRIVVDITSKPPATVEM+  
 Sbjct: 483 KISKRIVNEVDHVRIVVDITSKPPATVEMQ 513

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2143> which encodes the amino acid sequence <SEQ ID 2144>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.96 Transmembrane 228 - 244 ( 228 - 245)

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1383 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 RGD motif: 203-205

The protein has homology with the following sequences in the databases:

>GP: AAD15805 GB: AF058326 GMP synthase [Lactococcus lactis]  
 Identities = 411/511 (80%), Positives = 464/511 (90%), Gaps = 3/511 (0%)

50 Query: 10 VQKIIVLVDYGSQYNQLIARRIRFVGVSFELSKHKTATQELREINPIGIVLSCGPPNSVYD 69  
 ++KIIVLDYGSQYNQLIARRIRHGVFSEL SHK+TA+E+REINPIGI+LSCGPPNSVY +  
 Sbjct: 6 LEKIIVLVDYGSQYNQLIARRIRHGVFSELMSKVTAKIREINPIGIVLSCGPPNSVYD 65

55 Query: 70 NAPGIDPEIFELGIPILGICYQMQLITHKLGKVPAGQAGNREYQGSTLHLRETSKLF 129  
 +F IDPEIFELG+P+LGICYQMQL+++KLG V AG+ REYG + L L E S L F+  
 Sbjct: 66 GSFIDPEIFELGLFVLGICYQMQLMSYKLGWVEAGE---REYGVAPQLTEKSALFA 122

60 Query: 130 GTFQEQVLMSHGDAVTEIPESPHLVGSDNCPFAIENTEKNLYGIQPHPEVRHSVGN 189  
 GTP+ Q VLMSHGDT IPESPH+VG S + PFAA+ENTE+NLGYGIQPHPEVRHSV+G  
 Sbjct: 123 GTFPEQDVLMSHGDRVTALPEGPHVVCTSPNSPFAAVENTERNLYGIQPHPEVRHSVGT 182

65 Query: 150 DILKIFPAVNICARGDWSMDNFIDMETAKIRETVGDRKVLGLGGVDSSVGVLLQRAI 249  
 ++L+NFA++ICGA+G+WSM+NFIDM+I IRE VGD+KVLGLGGVDSSVGVLLQRAI  
 Sbjct: 183 EMLRNFALNICAGKWSMNFIDMQIKDIRKVGDKVLLGLGGVDSSVGVLLQRAI 242



|    |            |                                                               |     |
|----|------------|---------------------------------------------------------------|-----|
| 5  | Query: 250 | GEGLTCTIFVDHGLLRKDEBGDQVMGLGGKFGLNIRVDASKRFLDILLADVEPEKRRKII  | 309 |
|    | Sbjct: 243 | GEGLTCTIFVDHGLRKRK E DQVM LGGKFGLNII+VDA KRK+D L + DPR +KKII  | 302 |
| 10 | Query: 310 | GNFVYVFDDEASKLKGVDPLAQGTLYTDIIESGTETAQTIKSHHNVGSLPDMQFELIE    | 369 |
|    | Sbjct: 303 | GNFVYVFDDEANKLKGVDPLAQGTLYTDVIESGTDTAQTIKSHHNVGSLPDMQFQLIE    | 362 |
| 15 | Query: 370 | PLNTLKFDEVRALGIALGNPEIWMRQPPFGPGAIRVMAITHEKLETVRSDAILREE      | 429 |
|    | Sbjct: 363 | PLNTLKFDEVRALG LQMP+EIWMRQPPFGPGAIRV+G +TEEKLETVRSDAILREE     | 422 |
| 20 | Query: 430 | IAKAGLRDQWQYPTVNIIGVRSVGVNMGDKRTYDTYTAIRAITSIDGMTADPAQLPVDVLK | 489 |
|    | Sbjct: 423 | IA +GL+RDWQYPTVNI V+SVGVNMG DKRTYDTYTAIRAITSIDGMTADPAQLPVDLLQ | 482 |
| 25 | Query: 490 | KISTRIVNEVDHVNRIYDITSKPPATVENE                                | 520 |
|    | Sbjct: 483 | KISRIVNEVDHVNRIYDITSKPPATVENE                                 | 513 |

An alignment of the GAS and GBS proteins is shown below:

Identities = 487/520 (93%), Positives = 505/520 (96%)

|    |            |                                                                |     |
|----|------------|----------------------------------------------------------------|-----|
| 25 | Query: 1   | MTDISILNDIQKIIVLDYGSQYNQLIARRIREPGVFSELKSHKITADEIRIDNPIGIVLS   | 60  |
|    | Sbjct: 1   | MT+ISILND+QKIIVLDYGSQYNQLIARRIREPGVFSELKSHKITA E+R+INPIGIVLS   | 60  |
| 30 | Query: 61  | GGPNVSYADGAFGIDREIFELGIFILGICYGMQLITHKLGKVLNAGAGHREYQGSALR     | 120 |
|    | Sbjct: 61  | GGPNVSYADNAGFIDPEIFELGIFILGICYGMQLITHKLGKVVPAQAGNREYQGSTLR     | 120 |
| 35 | Query: 121 | LRSEBALFAGTPOBQLVLMSHGDVAITEPFGHVLNMGSDVDCPFAAMENITEKQFYGIQFHP | 180 |
|    | Sbjct: 121 | LR S LF+GTPQBLVLMSHGDVAITEPFGHVLNMGSDVDCPFAA+ENITEKLYGIQFHP    | 180 |
| 40 | Query: 181 | EVRHSVYGNILKNFAVNICGARGDWSMNFIDMEIAKIRETVADRKVLGLSGGVDSV       | 240 |
|    | Sbjct: 181 | EVRHSVYGNILKNFA+ICGARGDWSMNFIDMEIAKIRETVADRKVLGLSGGVDSV        | 240 |
| 45 | Query: 241 | VGVLQKAIGDQLCTCFVDHGLLRKDEBGDQVMGLGGKFGLNIRVDASKRFLDILLADVE    | 300 |
|    | Sbjct: 241 | VGVLQKAIGDQLCTCFVDHGLLRKDEBGDQVMGLGGKFGLNIRVDASKRFLDILLADVE    | 300 |
| 50 | Query: 301 | DPEKRRKIIGNEFVYVFDDEASKLKGVDPLAQGTLYTDIIESGTETAQTIKSHHNVGSLP   | 360 |
|    | Sbjct: 301 | DPEKRRKIIGNEFVYVFDDEASKLKGVDPLAQGTLYTDIIESGTETAQTIKSHHNVGSLP   | 360 |
| 55 | Query: 361 | EDMQFELIEPLNTLKFDEVRALGIALGNPEIWMRQPPFGPGAIRVMAITHEKLETVR      | 420 |
|    | Sbjct: 361 | EDMQFELIEPLNTLKFDEVRALGIALGNPEIWMRQPPFGPGAIRVMAITHEKLETVR      | 420 |
| 60 | Query: 421 | ESDAILREEIAKAGLRDQWQYPTVNIIGVRSVGVNMGDKRTYDTYTAIRAITSIDGMTADP  | 480 |
|    | Sbjct: 421 | ESDAILREEIAKAGLRDQWQYPTVNIIGVRSVGVNMGDKRTYDTYTAIRAITSIDGMTADP  | 480 |
| 65 | Query: 481 | AQLPVDVLKKIISTRIVNEVDHVNRIYDITSKPPATVENE                       | 520 |
|    | Sbjct: 481 | AQLPVDVLKKIISTRIVNEVDHVNRIYDITSKPPATVENE                       | 520 |

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 696**

A DNA sequence (GBSx0740) was identified in *S.agalactiae* <SEQ ID 2145> which encodes the amino acid sequence <SEQ ID 2146>. This protein is predicted to be branched chain amino acid ABC transporter, periplasmic amino acid-bind. Analysis of this protein sequence reveals the following:

```

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.0957 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9409> which encodes amino acid sequence <SEQ ID 9410> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

 >GP:AAD36211 GB:AB001771 branched chain amino acid ABC transporter,
 periplasmic amino acid-binding protein [Thermotoga maritima]
 Identities = 31/92 (33%), Positives = 51/92 (54%), Gaps = 4/92 (4%)

20 Query: 26 AKAFPHIHVYKAYGEEPMFSAISYDAVYMAKSAKGAQTSID---IKKALAKLKDFKQVT 82
 AK F + Y + YG+KP+ +AL YDA YM A S D I + + K +F G +
 Sbjct: 275 AKKFPVEYKYEKYGKPAALALALGYDA-YMVLLOAIERAGSPDREKTAIEIRKTRNPNFG 333

 Query: 83 GMSIDKRNINVVKSAIVVKLEDGKTSVNIIS 114
 G ++ID+N + +KS V +++G ++I+
25 Sbjct: 334 GIINIDEMDAIKSVVVMIVKNGSVDPEAVIN 365

```

No corresponding DNA sequence was identified in *S.pyogenes*.

30 SEQ ID 9410 (GBS660) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 8 & 9; MW 71.5kDa) + lane 10; MW 27kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 141 (lane 2; MW 46.5kDa) and in Figure 181 (lane 3; MW 46kDa).

GBS660-His was purified as shown in Figure 233, lane 5-6.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 697**

A DNA sequence (GBSx0741) was identified in *S.agalactiae* <SEQ ID 2147> which encodes the amino acid sequence <SEQ ID 2148>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

 INTEGRAL Likelihood = -10.61 Transmembrane 140 - 156 (129 - 158)
 INTEGRAL Likelihood = -9.55 Transmembrane 60 - 76 (53 - 80)
 INTEGRAL Likelihood = -7.59 Transmembrane 264 - 280 (257 - 285)
 INTEGRAL Likelihood = -5.79 Transmembrane 232 - 248 (219 - 251)
45 INTEGRAL Likelihood = -2.23 Transmembrane 190 - 206 (190 - 207)
 INTEGRAL Likelihood = -1.75 Transmembrane 90 - 106 (90 - 110)

 ----- Final Results -----
50 bacterial membrane --- Certainty=0.5246 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10059> which encodes amino acid sequence <SEQ ID 10060> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

5 >GP:AD36212 GB:AB001771 branched chain amino acid ABC transporter,
 permease protein [Thermotoga maritima]
 Identities = 140/295 (47%), Positives = 200/295 (67%), Gaps = 7/295 (2%)

10 Query: 2 LQQLNGLILGSIYALLAIGYTMVYGIILKINFAGHDYMMGAFMYSLINHLHLNFFLA 61
 LQ L NG++LG +TALA+AGYTMVYGI++LINFAGHD+ MMG + +Y L LN +
 Sbjct: 5 LQNLFNGIMGLGIALIAGITMVYGIILKINFAGHDVMMMGVYFAFATRLSLNLFPS 64

 Query: 62 LLIAMLGSAFLGVVITYLAUPELRKSTRIAALITAIGVSFLLEYGMVILNGADTRAPFQA 121
 ++R+LQ+A L3 +I+ +HY+PLR + RI+ALITAIGVSF LE V + GA ++F +
 Sbjct: 65 AIVAILGQALLSFLIDRVATKPLRNAPRISALITAIGVSFFLESLAVVVFPAIKPSFLKV 124

 Query: 122 IHTVKYHSGPITITNVQL-----IIIGIALMLMLTQFIVQKTKMKAMRALSVDSDAAQ 176
 +T+ ++ +++++ I ++++ L FIV +TK+G AMR+G+D
 Sbjct: 125 FKDRITLIKVLTVAGARIPILTFLVIFITAVILIVLFFIVYRTKIGAMRALSNDIPTFA 184

20 Query: 177 LMGINVRRTISFTFALGSALAGAGGVILGLYNSVQPIMGVTPGLKAFVAARVLGGIGIIP 236
 LMG+NV+ I FTFALGSALA A G++ + + +Y P MG FGLKAF+AV GIGIIP
 Sbjct: 185 LMGVNVDAVIGFTFALGSALAAAGSGIMHMKFENVHYPNGFMPLKAFIAAVFGSGISIP 244

25 Query: 237 GAAIGGVFGIGILETLATL--GVSDPRDGIVYAILILFLIRFAGILGKNIKKV 289
 GA +GG ++G++E A V +RD + IL+I L++P+G+LQK I EKV
 Sbjct: 245 GAVLGGVILGLIRFILAAYFPVWNYRDAPFAPILITLLVKEFSGLLGKTIKEV 299

```

There is also homology to SEQ ID 2150. A related sequence was also identified in GAS <SEQ ID 9171> which encodes the amino acid sequence <SEQ ID 9172>. Analysis of this protein sequence reveals the following:

```

 Possible site: 30
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.74 Transmembrane 196 - 212 (191 - 219)
35 INTEGRAL Likelihood = -12.42 Transmembrane 12 - 28 (5 - 36)
 INTEGRAL Likelihood = -7.22 Transmembrane 106 - 122 (102 - 126)
 INTEGRAL Likelihood = -4.78 Transmembrane 242 - 258 (240 - 260)
 INTEGRAL Likelihood = -2.50 Transmembrane 61 - 77 (60 - 77)
 INTEGRAL Likelihood = -2.34 Transmembrane 293 - 309 (291 - 309)
40 INTEGRAL Likelihood = -1.44 Transmembrane 139 - 155 (138 - 156)
 INTEGRAL Likelihood = -1.33 Transmembrane 317 - 333 (317 - 333)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.609 (Affirmative) < succ>
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

 Identities = 35/147 (23%), Positives = 71/147 (47%), Gaps = 6/147 (4%)

50 Query: 134 ITNVQLIIIGI--ALLMLTLQFIVQKTKMKAMRALSVDSDAAQLMGINVRNTISFTFA 191
 +TN I +GI A++ + + F+ KT +G +R++ ++ A++ G++ RTI +
 Sbjct: 197 LTNNSRINTGIFFAITATLWFLNKTTLGFEIRSVGLNPHASEYAGMSKRTIILSMI 256

55 Query: 192 LGSALAGAGGVLL--IGLYNSVQPIMGVTPGLKAFVAARVLGGIGIIRGAAGFVGIGILE 249
 + ALAG GGV+ +G + N + G ++L + G F+ G+L
 Sbjct: 257 ISGALAGLGGVVEGLGTFRNVFVQGSLLAVGEDGMVSLIAANSPL-GIFSSFLFGLVN 315

 Query: 250 TLATALGVSDPRDGIVYAILL-LFLI 275
 A ++ ++V + +IF +
60 Sbjct: 316 IGAPGMNTAGIPRLAKVVTASTIFFV 342

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 698

- 5 A DNA sequence (GBSx0742) was identified in *S. galactiae* <SEQ ID 2151> which encodes the amino acid sequence <SEQ ID 2152>. This protein is predicted to be branched chain amino acid ABC transporter, permease protein (livM). Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
10 INTEGRAL Likelihood = -8.76 Transmembrane 90 - 106 (84 - 113)
 INTEGRAL Likelihood = -8.23 Transmembrane 12 - 28 (5 - 33)
 INTEGRAL Likelihood = -8.17 Transmembrane 205 - 221 (200 - 224)
 INTEGRAL Likelihood = -7.86 Transmembrane 276 - 292 (273 - 300)
 INTEGRAL Likelihood = -6.32 Transmembrane 159 - 175 (154 - 176)
15 INTEGRAL Likelihood = -6.05 Transmembrane 236 - 252 (232 - 264)
 INTEGRAL Likelihood = -5.95 Transmembrane 42 - 58 (38 - 60)
 INTEGRAL Likelihood = -5.84 Transmembrane 120 - 136 (119 - 138)
 INTEGRAL Likelihood = -4.35 Transmembrane 255 - 271 (253 - 274)
 INTEGRAL Likelihood = -1.59 Transmembrane 66 - 82 (66 - 85)
20
----- Final Results -----
 bacterial membrane --- Certainty=0.4503 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
25

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AA036213 GB:AB001771 branched chain amino acid ABC transporter,
 permease protein [Thermotoga maritima]
Identities = 119/332 (35%), Positives = 191/332 (56%), Gaps = 33/332 (9%)
30 Query: 12 LAIVVLDYLLISVLISMGIFNLVHHQIIETIGINVLAVGLNLIVGCSQGFSLGHAGFMA 71
 L + V L + + + + + Y + + + + I I I + AV LNLI G + G PSLGHAGF +
 Sbjct: 16 LTVVFLIPMLLLYLADRYMDSYKLKRVRLIAIYGIAMVSLNLINGITGI PSLGHAGFTL 75
35 Query: 72 IGAYAVAIIGVKMP-----TYVGFLLAILVGLTVAGGIALGVGIPTLR 114
 IGAY + + + + + F A + G + + A A + G P L R
 Sbjct: 76 IGAYTASLLTSPQKAMSFIEFIVPWLNAHTDFFATVAGGVLAAVFALGNPVLRL 135
40 Query: 115 LKGDYLAATLGVAEIRILLVNGSDITNGAGGINGIPFPPTWSLVYGVAVVSLILAMNF 174
 L GDYLAATLG AE+IRI+ +N ITNG G+ GIP ++ YG V+++ +
 Sbjct: 136 LSGDYLAATLGLFAEIVRIIALNAISITNGPLGLKGIPEYSNIWVCYGNLFPVTLFMAEL 195
45 Query: 175 LRSPLGRNTIAIREDEIAAESMGVDITKVKVIVFVFGAILASIGSLQAGYVGTVMKPDF 234
 + S GR AIREDAAR+MG+ K + + FV GA A ++GSL A ++ T+ P+
 Sbjct: 196 VNSYGRALKAIREDRIARPMGIVNFKHQLLSFVIGAFAGVSGSLYAHMLITDPRRT 255
50 Query: 235 SF--MMSVNVLLIVVLGSLGSMGTFLVLAAILLGLNLMLQD-----YASVR 278
 + M++ VLI++VLGGLGS++G+++ A L +L L+D +R
 Sbjct: 256 TLGPMFLTYVLLIMIVLGLGLSISGLSGLAALFALLFSLHARDLSEPTTFPGIHVPGIKMR 315
55 Query: 279 MIYALALILIMIFRPSGLGTXKRLATSHLFR 310
 + + + + IL+MIF G+G +RLT ++L+R
 Sbjct: 316 ILVISAIPLVIMFPQRCINGRRLATVNNLYR 347

```

- 55 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 699**

A DNA sequence (GBSx0743) was identified in *S. agalactiae* <SEQ ID 2153> which encodes the amino acid sequence <SEQ ID 2154>. This protein is predicted to be branched chain amino acid ABC transporter, ATP-binding protein (livG). Analysis of this protein sequence reveals the following:

```

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2057(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:AAD36214 GB:AB001771 branched chain amino acid ABC transporter,
 ATP-binding protein (Thermotoga maritima)
 Identities = 136/271 (50%), Positives = 189/271 (69%), Gaps = 21/271 (7%)

 Query: 3 LLEVNLSKHFQGLTAVGDVSMKLNKGLIGLIGPNAGAKTTLFNLLTGVYLSKGTISI 62
 IL + +++ FGGL AV D + ++ +GEL+GLIGPNAGAKTT+FN++TG+Y P+KG I
20 Sbjct: 11 LLLLDHVTMQFQGLVAVDDPTNIBRGLVGLIGPNAGAKTTFNVTITGVYTPTKGRIVP 70

 Query: 63 DGKILNGRKPAKIASLGLGRTTFQNIIRLFKNMTVLNDLVGLSNHLSHPFASFLRLPK-- 120
 + + G + P + I LG+ RTFQNIIRLF +MTVL+NVLV +H LS+P A + +
25 Sbjct: 71 NDIDITGLRPYQITHLGIARTTFQNIIRLFSDMTVLNVLA-QHVLNPNFADRLILVKHK 129

 Query: 121 -----YTHSSKALKKALELLBIPGLKAYQDALANLPYGNQRRLBI 162
 Y KK + ++A +L++ GL+ A +LPTG+QR+LEI
30 Sbjct: 130 PRKHGRFWMFRAVTKITGYLKEKEMVERANDLIKRVGLEKVMYTKASSLPYGEQRRLBI 189

 Query: 163 VVALATEPKILFLDEPAAGNPNQETAEITQLIQKSDPDITMLHLEHDMVAVQVTERI 222
 RALATEPK++ LDSPAAGNPN+ET +L + I QI+ DP++P++LLEHDM +VM + ERI
35 Sbjct: 190 ARALATEPKILFLDEPAAGNPNKZTEDLMFTIKQIRKDNLITVLLIHEDMKVWMSICERI 249

 Query: 223 YVLEVGRLLAHOTPEETIKNNKQVIRAYLGE 253
 V++YGR+IA GTP+ET+H+ RVTEAYIG E
 Sbjct: 250 IVMYGRLLAHOTPEETIKNNKQVIRAYLGE 280

```

There is also homology to SEQ ID 644.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 700**

A DNA sequence (GBSx0744) was identified in *S. agalactiae* <SEQ ID 2155> which encodes the amino acid sequence <SEQ ID 2156>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.2216(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

55 >GP:CA852068 GB:AL109732 putative branched chain amino acid
 transport ATP-binding protein (Streptomyces coelicolor
 A3(2))
 Identities = 136/233 (58%), Positives = 181/233 (77%)

```

-800-

Query: 3 MLAVENLSIHVGVIQAVNDVSEFVNQCEVVTLIGANGAGKTSILRTISGLVRPESQ3SIF 62  
 +L+VE+L + YG I+AV +SF+V+ GEVVTLLIG NGMGKT+ LRT+8GL++P G I F  
 Sbjct: 4 LLEVEDLRVAYGKIEAVKGISPKVDAGEVVTLLIGTNGAGKITTTLRTLSGLLKPVGGQIRF 63

Query: 63 NGKPIHKLAARKIVNGLAQVPEGRHV75SLVMENLAWGAPLQKDRBQNRMLKKVDFR 122  
 GK + K+ A +IV GLA PGRH+F +++ +NL+ +GAFI+ DR +K +++ +D  
 Sbjct: 64 GGRSLKKVPAPHQIVSLGLAHSF63RHIFFRMITIEDNLRLGAPLRSRDRGIEKDQRAYDL 123

Query: 123 FFLREERKNDAATLSGGSQMLAMGRALMERPKILLDDPEWGLAPITQIEFMYLIEDI 182  
 PP L ER+ Q A TLSGGSQMLAMGRALMS+PKILL+LDEPSWGLAPIT +Q+I I ++  
 Sbjct: 124 FFLIGERRRQAAGTLSGGSQMLAMGRALMSQPKILL+LDEPSWGLSPITMQKIMATIAEL 183

Query: 183 KKGQTVTLVEQNNAKALTADKAYVLERGVVLSGTGKELLVSDQVRKAYLG 235  
 K QGTT+LLVEQNA AL++AD +V+E G +VLSG+G+LL + VRKAYLG  
 Sbjct: 184 KSGQTVTLVEQNNAKALSLADHGHVMEVGVVLSGSGQDLLHDSVRKAYLG 236

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 20 Example 701

A DNA sequence (GBSx0745) was identified in *S.agalactiae* <SEQ ID 2159> which encodes the amino acid sequence <SEQ ID 2160>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0415 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP: AAD36216 GB: AE001771 conserved hypothetical protein [Thermotoga maritima]  
 Identities = 72/166 (43%), Positives = 116/166 (69%), Gaps = 2/166 (1%)

Query: 1 NPVKDFMTKLYVSEDTTVAEADLLREHHLRLFPVENDQVGLVTEGTMAEAPSKA 60  
 N VKDFMT+ + ++P+T+ +EA L++++ +RL V+H+++VG+VTE + A PSKA  
 Sbjct: 1 MLVKDFMTRNPITTAPESTPSSEALKMKQNKIKRLVWVKNEKIVGVTEKDLLXASPSGA 60

Query: 61 TSLSIYB+VYLLNKTKIRDTMIKDITVVSQVASEDAIYLMMSRKIGVLPVDN+GQLYG 119  
 T+L+I+e++YLL+K KI +IM KD+VTV++ +EDA +M + I LPUVD+ G+L G  
 Sbjct: 61 TTTNIWELHYLLSKLKTIEINTKDVTVMHENTPEDIAARIMEKDISGLFPVVDACRLVG 120

Query: 120 IVTDRDVPKAFLEIAGYGC+SYRLVILADRGIVGLSKVLNRLSSA 164  
 I+T D+PK F+EI G +E + R + + G L +V R+ A  
 Sbjct: 121 ITTQDIPKVPVEIPGTGREGTIRYTMEMPDKPGLLEVAKRIVYA 166

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 702

A DNA sequence (GBSx0746) was identified in *S.agalactiae* <SEQ ID 2163> which encodes the amino acid sequence <SEQ ID 2164>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5585 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-801-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 703

A DNA sequence (GBSx0747) was identified in *S.agalactiae* <SEQ ID 2165> which encodes the amino acid sequence <SEQ ID 2166>. This protein is predicted to be a transposase. Analysis of this protein  
10 sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.65 Transmembrane 53 - 69 (53 - 70)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

- 20 The protein has homology with the following sequences in the GENPEPT database:

```
>GF:AA85003 GB:U28972 SpV1 ORF3; putative transposase [Spiriplasma citril]
Identities = 49/154 (31%), Positives = 80/154 (51%), Gaps = 11/154 (7%)

25 Query: 39 WLEMDTVIGRIGGKVLITFNVAFCNFI FAKLNDKTAIETAKHIQ--VIRKTLYDNKRF 96
 WLEMDTV+G+ ++L FA +++ TA E K + +IK L +
 Sbjct: 174 WLEMDTVVGKDHKSAILVLVBQLSKKYFAIKLENHTAREVKKFKDIIIRNNLIGIKG- 232

Query: 97 FELFFVILTDNGGEFARVDDIZIDVCGSQCLFFCDPNRSDQKARIEKNHTLVRDILPKGT 156
 I+TD G EF++ ++EI ++Q++PCD QK IE ++ +R FNGT
30 Sbjct: 233 -----IITDRGKEFSKWRMEI--FAETQVYFCDAGSPQQKPLIEYNNSELARHWFFNGT 284

Query: 157 SFDNLTVQEDINLALSHINSVKRQALANGKTAVELF 190
 F+ ++Q+ I+ ++ IN R LN ++ E+F
35 Sbjct: 285 DFNKVSQKQIDWVNVINDKLRPLCLAWLISKENF 318
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 704

- 40 A DNA sequence (GBSx0748) was identified in *S.agalactiae* <SEQ ID 2167> which encodes the amino acid sequence <SEQ ID 2168>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3116(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

- 50 A related GBS nucleic acid sequence <SEQ ID 10055> which encodes amino acid sequence <SEQ ID 10056> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 5 Example 705

A DNA sequence (GBSx0749) was identified in *S.agalactiae* <SEQ ID 2169> which encodes the amino acid sequence <SEQ ID 2170>. This protein is predicted to be thymidylate kinase (tmk). Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1876 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10053> which encodes amino acid sequence <SEQ ID 10054> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB03761 GB:AP001507 thymidylate kinase [Bacillus halodurans]  
 Identities = 112/210 (53%), Positives = 148/210 (70%), Gaps = 1/210 (0%)  
 Query: 17 MKKGLMISFEGPDGAGKTTVLEAVLPILREKLISQDILTTREPGGVITISEIRHIILDVKH 76  
           M KG I+ EG +GAGKT+ L+A+ +LRE ++ TREPGG+ I+E+IR IILEV H  
 25 Sbjct: 1 MTKGCFITVEGSGAGKTSALDAIEENLEN-GLSVVTRTEPGGIPIAPQIRSIILVDH 59  
 Query: 77 TOMDKKTELLLYMAARRQHLVEKVLPALESGKIVIMDRFIDSSVAYQSGRGLDKSHKN 136  
           T+MD +TE LLY AARRQHLVEKVLPALE G +VL DRFIDSS+AYQG RG+ I  
 30 Sbjct: 60 TRMDPRTEALLYAARRQHLVEKVLPALEAGHVLCORFIDSSLAYQGYARGIGFEDILA 119  
 Query: 137 LNDYATDSHKPDILTYFDVPSEVGLEIKQSVQREVNRDLQDLMDHQRVVRQGYLELADS 196  
           +N+ +A + PDLTL F V +VGL RI + RE NRDL E L HQ+V+ +GY ++  
 35 Sbjct: 120 INEFAIEGRYPDLTLFVDPDVGSLRIHRDQSRBQNRDLQRAITPHQKVBKEGYERIVT 179  
 Query: 197 EPNRIVITIDASQQLDEVIATFPIILDRIN 226  
           P R+V IDA+Q D+V+A+ +I R++  
 Sbjct: 180 YPERVVEIDANQSPDQNVADAVRMKQRIS 209

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2171> which encodes the amino acid sequence <SEQ ID 2172>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.75 Transmembrane 215 - 231 ( 215 - 231)  
 45 ----- Final Results -----  
           bacterial membrane --- Certainty=0.1298 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the databases:

>GP:BAB03761 GB:AP001507 thymidylate kinase [Bacillus halodurans]  
 Identities = 109/205 (53%), Positives = 148/205 (72%), Gaps = 1/205 (0%)  
 Query: 22 MITGKLTIVSGPDGAGKTTVLEQLIPLLKQKVAQDILTTREPGGVATSEIHREILIDIN 81



-803-

M G ITVBG +GAGKT+ L+ + +L++ ++ TRBPGG+ I+E IR +ILD+H  
 Sbjct: 1 MTGCFITVGGGAGKTALDAIEMLRN-GLSVVTRBPGGIPABQIRSIILVDH 59  
 Query: 82 TAMDPKTELLLYAARQHIVKVLPALEAGQLVFIDRFIDSSVAYQAGRGLIKADICW 141  
 T MDP+TE LLY AARQHIVKVLPALEAG +V DRFDSS+AYQG R3+ DI  
 Sbjct: 60 TRMDPTEALLLYAARQHIVKVLPALEAGHVLCDFIDSSLAYQGYARGIGFEDILA 119  
 Query: 142 INEFATDGLSPDLTYFDVPSEIGLARINANQOREVNRLOLSTETIHRVKGYLALAKE 201  
 +NEFA +G PDLT P V +GL+RI+ +Q RE NRLE E+ HQ+V+GY + +  
 Sbjct: 120 INEFATBGRYPDLTYLFDVDFVGLSKIRHDCREONRLDQEALTTHKVRGYSRIVET 179  
 Query: 202 HPRKRVITIDATKPLKEVVSVALEHV 226  
 +P+R+V IDA + +VV+ A+ +  
 Sbjct: 160 YFRRVVEIDANQSFQVVAIDVMI 204  
 Query: 4 FDRIVVIINRGKTRMKGLMISFSGDQAGKTTVLEAVLPILREKLSQDILITREPGGVTI 63  
 PD+I ++ ++G M G +I+ BGFQDQAGKTTVLE +PLL++K++QDILITREPGGV TI  
 Sbjct: 9 FDKIELLSBGNKHITGKLTITVBGPDQAGKTTVLEQLIPLLKQVQDILITREPGGVAI 68  
 Query: 64 SEHIRHILDVHGTOMDKKTELLLYMAARQHIVKVLPALEBEGKIVLMDRFIDSSVAYQ 123  
 SE IR +ILD+ HT ND KTELLLY+AARQHIVKVLPALE G++V +DRFDSSVAYQ  
 Sbjct: 69 SEHIRELILDIHNTMDPKTELLLYAARQHIVKVLPALEAGQLVFIDRFIDSSVAYQ 128  
 Query: 124 GSGRGLDKSHIKWINDYATDSHPDLTYFDVPSEVGLERIQSVQOREVNRLOLQDMH 183  
 G+GRGL K+ I+WIN++ATD +PDLTLYFDVPSE+GL RI +QREVNRLDLE +++H  
 Sbjct: 129 GAGRGLIKADICWINDEFATDGLSPDLTYFDVPSEIGLARINANQOREVNRLOLSTETI 188  
 Query: 184 QRVRQGLYELADSENRIVITIDASQQLDEVIAETPSIIL 222  
 QRVR+GXL LA P RIVTIDA++ L EV++ +L  
 Sbjct: 189 QRVRKGYLALAKEHPKRVITIDATKPLKEVVSVALEHVL 227

An alignment of the GAS and GBS proteins is shown below:

Identities = 145/219 (66%), Positives = 181/219 (82%)

Query: 4 FDRIVVIINRGKTRMKGLMISFSGDQAGKTTVLEAVLPILREKLSQDILITREPGGVTI 63  
 PD+I ++ ++G M G +I+ BGFQDQAGKTTVLE +PLL++K++QDILITREPGGV TI  
 Sbjct: 9 FDKIELLSBGNKHITGKLTITVBGPDQAGKTTVLEQLIPLLKQVQDILITREPGGVAI 68  
 Query: 64 SEHIRHILDVHGTOMDKKTELLLYMAARQHIVKVLPALEBEGKIVLMDRFIDSSVAYQ 123  
 SE IR +ILD+ HT ND KTELLLY+AARQHIVKVLPALE G++V +DRFDSSVAYQ  
 Sbjct: 69 SEHIRELILDIHNTMDPKTELLLYAARQHIVKVLPALEAGQLVFIDRFIDSSVAYQ 128  
 Query: 124 GSGRGLDKSHIKWINDYATDSHPDLTYFDVPSEVGLERIQSVQOREVNRLOLQDMH 183  
 G+GRGL K+ I+WIN++ATD +PDLTLYFDVPSE+GL RI +QREVNRLDLE +++H  
 Sbjct: 129 GAGRGLIKADICWINDEFATDGLSPDLTYFDVPSEIGLARINANQOREVNRLOLSTETI 188  
 Query: 184 QRVRQGLYELADSENRIVITIDASQQLDEVIAETPSIIL 222  
 QRVR+GXL LA P RIVTIDA++ L EV++ +L  
 Sbjct: 189 QRVRKGYLALAKEHPKRVITIDATKPLKEVVSVALEHVL 227  
 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 706

A DNA sequence (GBSx0750) was identified in *S. agalactiae* <SEQ ID 2173> which encodes the amino acid sequence <SEQ ID 2174>. This protein is predicted to be DNA polymerase III delta' subunit (dnaXX).

Analysis of this protein sequence reveals the following:

Possible site: 26  
 >> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2603 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB03763 GB:AP001507 DNA polymerase III delta' subunit [Bacillus halodurans]  
 Identities = 78/189 (41%), Positives = 113/189 (59%), Gaps = 3/189 (1%)  
 Query: 2 DLKRTQPKLEKFNITLQSDRNSHAYLFSGNFAS--LDWALYLAQSOFCEKRSGLGPCQR 59  
 +L + QP + L R++RAY+P GN + MAL+LA+S FC +R PCQ  
 Sbjct: 5 NLAKIQPFVATMLKNSLAKRLAHAYIPDGNRGTKKRWALHLAKSFFPCAQAGVEPCQT 64  
 Query: 60 CRACRLTANGBFSVDVKIIEPQGLIKETIKELTKDFSRGFSBKGQVFIKDCREMIYN 119  
 C+ C+ I +G DV IEP QG IK ++ L K+FS G E +V+I+ +RN +  
 Sbjct: 65 CKECKRIEHKNIPDVHFIETDQSQSIKKQVEHLQKRFPSYRGMEGAKKYVIIVHADRIITS 124

-804-

Query: 120 AANSLKPFIEEPQSSSVILLINDENNVLPTIKSRTOIFRF-PKQDMLVHQABQGLLK 178  
 AANSLKPFIEEP + + ILGT N+LPTIKER+Q+ P P + + E+ G+ +  
 Sbjct: 125 AANSLKPFIEPLADTVAILLTEQLQMLMPTIKSRQVLSFAPLEVQAFKLIEEGISE 184

5 Query: 179 SQASLLAQV 187  
 S ++LLA +  
 Sbjct: 185 SVSNLLASL 193

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2175> which encodes the amino acid sequence <SEQ ID 2176>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2685(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 151/290 (52%), Positives = 213/290 (73%), Gaps = 3/290 (1%)

Query: 1 MDLKRTPQKLEKNTIILQSDRMHAYLPSGNEFASLDNALYLAQSQFCEKQSGLPQCEC 60  
 MDL + P + + F TIL+ DR++HAYLPSG+FA+ +MAL+LA+ FCE+++ PC C  
 Sbjct: 1 MDLQAENAVNYQAFQTLIKKDRLNHAYLPSGDFANESHALFLAKVIFCEQKQKQTPGHG 60

25 Query: 61 RACRLIANGEPFSDVKIIEPQGLIKTETIKELIKDFSRSGFBGKSQVFIHKDCBKMHYNA 120  
 R+CLII G+P+DV ++EP QQ+IKT+ +KE+ +FS++G+E K QVFIHKDC+KMH+NA  
 Sbjct: 61 RSQQLISQGFADVTIVLEPTGQVITKTDVVKENWAFNSOTSYENKROVFIHKDCDKNHNA 120

30 Query: 121 AANSLKPFIEEPQSSSVILLINDENNVLPTIKSRTOIFRF-PKQDMLVHQABQGLLK 180  
 AANSLK+IEEPQ +Y+ LLTND+N VLPTIKSRTO+P+FPK L A++ GLL Q  
 Sbjct: 121 AANSLKYIEEPQGEAYIFLLTNDNDKVLPTIKSRTOVPOPPHNAVLYQLAQEKGILLNHQ 180

35 Query: 181 ASLLAQVADDPKHLEILLTKKGLLDYINLSQFVTTIANDROTAYLEVSRVLTQVVDKND 240  
 A L+A++R+ HLE LL KLL+ + +++FV+ KD+ RYL ++RL +K +  
 Sbjct: 181 AKLVAKLATNTSHLERLLQTSKLELLITQERFVSVINLQDLQAYILANRLVQLATEKEE 240

Query: 241 QAFVFOKLTIMLAKE---GGLYDIENTYRAQNMKNVSPNSLEYMVLS 287  
 Q V LT++LA+E L LE Y+R+ MW+SNV+PQN+LEYMV+S  
 40 Sbjct: 241 QDLVLTLLTLLARERACTFLTQLEAVYQARLNMQSNVNPQNTLEYMVMS 290

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 707

45 A DNA sequence (GBSx0751) was identified in *S.agalactiae* <SEQ ID 2177> which encodes the amino acid sequence <SEQ ID 2178>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2016(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAR03765 GB:AP001507 unknown conserved protein in B. subtilis  
 [Bacillus halodurans]  
 Identities = 45/116 (38%), Positives = 62/116 (52%), Gaps = 8/116 (6%)

-805-

Query: 1 MDKKDLFDAPDFDSQNLVLGLSEIETMKQIQKLLSEMTVLRIENGKLRERLSVIEAET- 59  
 M+KK +F + + E+ +K+Q+ L+EEN L IEN LRERL E E  
 Sbjct: 1 MNKKAIPFQVSQKLERIGELHRELGLGKQLAYLIEENHPLTIEENHRLERLGEFLEET 60

5 Query: 60 ---STAVKNSK----QGRELLEGIYNDGFHICNTFYQORRENDEBCAPCIELLYRD 108  
 E K K +G + L +Y +GFHICNT YG R+N E+C FC+ L +D  
 Sbjct: 61 EEKQVTKERKPFVVGEGYDNLARLQEGEHIINTYHGLAKNGEDCLFCLSPFLAQD 116

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2179> which encodes the amino acid sequence <SEQ ID 2180>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0700(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 75/107 (70%), Positives = 89/107 (83%), Gaps = 1/107 (0%)

Query: 1 MDKKDLFDAPDFDSQNLVLGLSEIETMKQIQKLLSEMTVLRIENGKLRERLSVIEAETE 60  
 ++KK+LFDAPDFDSQNL+V L+EIE MKRQ+Q L+SENT+LR+EN KLRERLS +E ET  
 Sbjct: 1 VNKKELFDAPDFDSQNLMTVLAEIEMKQVQSLVEENTILALENTKLRERLSHLEKET- 59

25 Query: 61 TAVKNSKQGRELLEGIYNDGFHICNTFYQORRENDEBCAPCIELLYR 107  
 A SXQ ++ LEGIY++GFHICN FYQORRENDEBC FC ELL R  
 Sbjct: 60 VAINPSKQRKDLEGIYDGFHICNFFYQORRENDEBCMPCELLDR 106

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 708

A DNA sequence (GBSx0752) was identified in *S.galactiae* <SEQ ID 2181> which encodes the amino acid sequence <SEQ ID 2182>. Analysis of this protein sequence reveals the following:

35 Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.28 Transmembrane 119 - 135 ( 119 - 135)

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1510(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10051> which encodes amino acid sequence <SEQ ID 10052> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB03768 GB:AP001507 unknown conserved protein [Bacillus halodurans]  
 Identities = 138/287 (48%), Positives = 189/287 (65%), Gaps = 2/287 (0%)

50 Query: 4 MQVQSFKNIIHYGLYLVPTPIGNLDMTPAIRILREYDFICAEPTRTNGLLLKHFDI 63  
 M+ Q+S++ GTLYLV TPIGNL+D+TPAIR L+E D I AEPTR T LL HFDI  
 Sbjct: 1 MKTQSSYQQRDDKGTLYLVATPIGNLDEVTTPAIRILKEADQIAAEITRQTKLLNHFPI 60

55 Query: 64 TTKQISFHEHNAYDKISGLIDLLKRGKSLACVSDRGMPISDPGHLVKAIEGDIPTVS 123  
 TK +S+HEHN LID L EG++A VSDRGMP+ISDPG++LV +AI+ I V+  
 Sbjct: 61 ATKLVSYHEHNKETMKRRLIDLLKRGKSLACVSDRGMPISDPGHLVWSAIEGLAVIP 120

-806-

Query: 124 IPGASAGITALIASGLAPQPHIFYGFLPRKKGQITTFETKQDYPTQIFYESPFRVSOT 183  
 IGR+A +TALIASGL + F GFLPR+K Q+ E + T IFYESP R+ DT  
 Sbjct: 121 IPGANAAVITALIASGLPTESQFIGFLPRQKQKQALESTKPKATLIFYESPHRLKDT 180

5 Query: 184 LKHKMEIYGDQVVLVRELTKLYEYORGTISQLLHIEKVPKLGECILIVDGKRDTERV 243  
 L M I G+R V + RELTK YEE+ RGT+ + + + +KGE +IV+G +  
 Sbjct: 181 LDMMLILGNRHVSI CRSLTKTYREFLRGTLESAVHWAREATIKGEFLIVBNGEKEVSP 240

10 Query: 244 KDS--SQDDPLVLKVEYIANKDKTQAIKKVAKFNINRQELVYASH 238  
 ++ P+ V+ YIA G++ +AIK+VA + +++++Y +H  
 Sbjct: 241 EEVWSESLSPVQHVHYIALGFRSKEAIKQVATDRGVPKRDYINHYH 287

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2183> which encodes the amino acid sequence <SEQ ID 2184>. Analysis of this protein sequence reveals the following:

15 Possible site: 35  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.09 Transmembrane 116 - 132 ( 116 - 134)

20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2635 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:BA03768 GB:AP001507 unknown conserved protein [Bacillus halodurans]  
 Identities = 139/287 (48%), Positives = 189/287 (65%), Gaps = 2/287 (0%)

30 Query: 1 MQVQKSPDKKTKSGTLYLVPTPIGNLDMTFRAVATLKEVDPICAEDTRNTCLLKHEDI 60  
 M+ Q+G++ + GILYLV TPIGNL+D+TFRA+ TLKE D I AEDTR T LL HEDI  
 Sbjct: 1 MKTQSQYQRRDKGTYLYVATPIGNLSDVTFRAITLKEADQIAEDTRQTKLLAHEDI 60

35 Query: 61 ATKQISFHEHNAYEKIPDLIDLLISGRSLAQVSDAGMPISDPGHDLVKAADSDIAVVA 120  
 ATK +SHAEHN LID LI GR+A VSDAGMP+ISDPG++LV +AI IAV+  
 Sbjct: 61 ATKLVSYHEHNKETMGKRLIDDLISGRSLAQVSDAGMPAISDPGYELVVAIAKEGIAVPI 120

40 Query: 121 LFGASAGITALIASGLAPQPHVFGFLPRKAGQKAFEDKHYPETQMFYESPYRKDT 180  
 +PGA+A +TALIASGL + F GFLPR+ Q++ E+ T +FYESP+R+KDT  
 Sbjct: 121 IPGANAAVITALIASGLPTESQFIGFLPRQKQKQALESTKPKATLIFYESPHRLKDT 180

45 Query: 181 LTNMLACYGDQVVLVRELTKLFEEYORGSISEILSYLSETPKLGECILIVA--GAQADS 238  
 L +ML G+R V + RELTK +RE+ RG++ E + + E +KGE LIV G++  
 Sbjct: 181 LDMMLILGNRHVSI CRSLTKTYREFLRGTLESAVHWAREATIKGEFLIVBNGEKEVSP 240

45 Query: 239 EVELTPADVLSLVOKEIQAGAKFNQAIKIAQVYNRQELVQOOF 285  
 E + V V+ I G + +AIK +A V++++Y +H  
 Sbjct: 241 EEVWSESLSPVQHVHYIALGFRSKEAIKQVATDRGVPKRDYINHYH 287

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 208/287 (72%), Positives = 238/287 (82%)

Query: 4 MQVQKSPKSNHLYGTLYLVPTPIGNLDMTFRAVATLKEVDPICAEDTRNTCLLKHEDI 63  
 MQVQKSPK GILYLVPTPIGNL DMTPRA+ L+EDVPICAEDTRNTCLLKHEDI  
 Sbjct: 1 MQVQKSPDKKTKSGTLYLVPTPIGNLDMTFRAVATLKEVDPICAEDTRNTCLLKHEDI 60

55 Query: 64 TTKQISFHEHNAYDKISGLIDLLKSGSLAQVSDAGMPISDPGHDLVKAADSDIAVVA 123  
 TKQISFHEHNAY+KI LIDLL G+SLAQVSDAGMPISDPGHDLVKA+ DI VV+  
 Sbjct: 61 ATKQISFHEHNAYEKIPDLIDLLISGRSLAQVSDAGMPISDPGHDLVKAADSDIAVVA 120

60 Query: 124 IPGASAGITALIASGLAPQPHIFYGFLPRKKGQITTFETKQDYPTQIFYESPFRVSOT 183  
 +PGASAGITALIASGLAPQPH+FYGFLPRK GQQ FFE K YPTQ+FYESP+R+ DT  
 Sbjct: 121 LFGASAGITALIASGLAPQPHVFGFLPRKAGQKAFEDKHYPETQMFYESPYRKDT 180

Query: 184 LKHKMEIYGDQVVLVRELTKLYEYORGTISQLLHIEKVPKLGECILIVDGKRDTERV 243  
 L +M YGDQVVLVRELTKL+EEYORGT+IS++L +E+ ELKGECL+IV G+ V

-807-

Sbjct: 181 LTNMLACYGDROVVLVRELTKLFESYORGSSISEILSYLBETPLKGECLLIVAGAOADSEV 240

Query: 244 KDSOQOPIVLVKEYIANGDKTNOAIIKKVAKEFNINRCKLYASEHDL 290

5 sbjct: 241 ELTADVDLVSLVOKEIOAGAKPNOAIKTIAXAYOVNROELYOOFHDL 287

A related GBS gene <SEQ ID 8643> and protein <SEQ ID 8644> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 10
MoG: Discrim Score: -6.92
GVH: Signal Score (-7.5): -9.26
 Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -1.28 threshold: 0.0
TWINS Likelihood = -2.26 Transmembrane 118 - 134 (118 - 134)
PERIPHERAL Likelihood = 6.89 32
modified ALOM scores: -0.76

```

\*\*\* Reasoning Step: 3

```

----- Final Results -----
 bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

EF002663.1 (200 - 1164 of 1470)
EGAD137863|BS003612(2 - 289 of 292) hypothetical protein 33.0 kd protein in xpac-abrb intergenic
region {Bacillus subtilis} OMNI|T010BS0044 conserved hypothetical protein
SP|P37544|YABC_BACSU HYPOTHETICAL 33.0 KDA PROTEIN IN XPAC-ABRR INTERGENIC REGION.
GP#67425|dbj|BA05271.1|D26185 unknown {Bacillus subtilis}
GP#263203|emb|K1812.1|J299104 similar to hypothetical proteins {Bacillus subtilis}
FIR|S66065|S66065 conserved hypothetical protein yabc - Bacillus subtilis
%Match = 24.5
%Identity = 45.8 %Similarity = 65.7
Matches = 131 Mismatches = 97 Conservative Sub.s = 57

```

CSTH\*KW\*TS\*A.SERY\*SRNRNCS\*KF\*TRKRI~~T~~RHLQ\*WLSHL\*YFLMSTS\*K\*R~~KRMC~~FLY\*III\*RLMENQVQSFK  
:  
  
MLPROMSFN

363            393            423            453            483            513            543            573  
SNHYGTLTYLPTPIGNLDMDTFRAIRILREVDFICABDTRNTGLLLKHFDITTKQISFHEHNAYDKISGLIDLLKSGKS  
| | | | | | | | | | | | | | | | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
GKSDMGLYLTYLPTPIGNLEMDTFRAIDT.KSVDAIAASDTORTKKLCVYIELETPLVS YHEHNKESGHKI IEWLKSGKN  
20            30            40            50            60            70            80

603            633            663            693            723            753            783            813  
IAQVSDAGMPISDPGRDLVKAAIEGDI PVVISPGASGITIALIASGLAPPHIFYGFLPRKKGOOITTFETKQOYPETQ  
| | | | | : | | | : | | : | | : | | | : | | | : | | : | | : | |  
IALVSDAGLPTISDPGARIVKDFDTIGGVVPLPGANNAULTI ASGI VPPFFYGFGLNROKKEKKELEALKKRQSTI

100            110            120            130            140            150            160

[illegible]

```

1074 1104 1134 1164 1194 1224 1254 1284
LVL---VKEYTANGDKTQAIKKVAKBNLNRQIYASFHDL*VII*KGCRKTIQWPFITSLAIGIKK*UTSNFKLFIEN
 | | | | | : | : | | | : | : | : | : | : |
LTAKEHVHYHISKGATSKALKKAADVDRNVKPREVVDYAHITQ
 260 280 290

```

-808-

SEQ ID 8644 (GBS343) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 11; MW 35.4kDa).

The GBS343-His fusion product was purified (Figure 215, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 277), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 709

A DNA sequence (GBS0753) was identified in *S. agalactiae* <SEQ ID 2185> which encodes the amino acid sequence <SEQ ID 2186>. This protein is predicted to be bA483F11.3 (cutC). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2568 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CA08199 GB:AL133353 bA483F11.3 (CGI-32 protein) (Homo sapiens)
Identities = 79/203 (38%), Positives = 116/203 (56%), Gaps = 7/203 (3%)

Query: 3 LRFPCAENITDITRLDKAIIISRVELCDNLAVGGTTPSYGVIKENQYLHEKGISVAVMIR 62
L E C + + + + + R + E L C L + G G T T P S G V + + Q + I V V M I R
Sbjct: 27 LNEVCVDSVESAVNAERGGADRIELCSGLSEGGTTPSMGVLVQVKSQVQ---IPFVFMIR 83

Query: 63 PRGGNFVYNDLEIRIMEEDILRAVELSDALVGLITSNHIDTEAIBQLLEPATQGLPLW 122
PRGG+P+Y+D E+ +M+ DI A +D LV G LT + HID E L+ + LP+
Sbjct: 84 PRGGDPLYSREIVMKADIRLAKLYGADGLVFGALTEDGHIDKELCMSLMAICRLPLPT 143

Query: 123 FHMADFVIPSQDKSIDQLVALGFTIRLLHGSSNGEPIIENIKHIALVEYANRIEIM 182
FH AFD + + D + + + L+ LGF R+L G + +E + IK L+E A RI +M
Sbjct: 144 FHRAFDW---HDPMALETLLTGLFERVLTSQCDSS--ALEGLPLIKRLIEQAQRIVVM 199

Query: 183 VGGGVTAENYVYICQETGVKQAH 205
GGG+T N Q I + +G + H
Sbjct: 200 PGGGITRNLQRILEGSGATEFH 222

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2187> which encodes the amino acid sequence <SEQ ID 2188>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2372 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 143/208 (68%), Positives = 168/208 (80%)

Query: 2 ILRFPCAENITDITRLDKAIIISRVELCDNLAVGGTTPSYGVIKENQYLHEKGISVAVMIR 61
+++RFPCEANIT L L D ISRVELCDNLAVGGTTPSYGVIKR Q LH+K ISVA MI

```

-809-

Sbjct: 1 MIKEPCAENLTLLPTLDAGQISRVELCDNLAVGGTTPSYGVIKACQLLEDRKISVATMI 60

Query: 62 RPRGGNFVYNDLELRIMEDILRANVELSDALVLGLITNNHIDTAIBQLLEPATQGLPL 121  
RPRGG+FYNDLELR+ MEEDIL+AVE SDALVLG+LT+ N +DT+AIQLLEPATQGLPL

5 Sbjct: 61 RPRGGDFVYNDLELRAMEEDILKAVEAGSDALVLGLLATTENQLVTAIBQLLEPATQGLPL 120

Query: 122 VFHMAFDVLPKSDQKKSIDQLVALGPTRIILHSGSSNGRPTIIRNIKIKALVEYANNRIEI 181  
VFHMAFD IP Q + + + DQL+ GP R+L HGS PI +N++ +K+LV YAN RIEI

10 Sbjct: 121 VFHMAFDRIPTDQHQAOLDQLIDYGFVRVLTHGSPRATPTTDNVEQLKSLVYANKRIEI 180

Query: 182 MVGGGVFARNYQVICOBTGVQKAGHRI 209  
M+GGG+TREN Q + Q TG HOT+I

Sbjct: 181 MIGGGITABNCQSLGQLTGTAVHGTKI 208

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 710**

A DNA sequence (GBSx0754) was identified in *S. agalactiae* <SEQ ID 2189> which encodes the amino acid sequence <SEQ ID 2190>. Analysis of this protein sequence reveals the following:

20 Possible site: 23  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1216 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:BAA12206 GB:D84061 phosphoserine aminotransferase [Spinacia  
oleracea]

Identities = 65/109 (59%), Positives = 79/109 (71%), Gaps = 1/109 (0%)

Query: 3 IYNFAGFAVLKPKPVLVKAQSELLNVQGSMSVLEVSHRKRFDDIIGAERYLRDLMI 62  
++NF+AGFAVL+ VL KQSELLN++GS MSV+E+SHR KRF II AE LR L+ I

35 Sbjct: 69 VFNFAAGFAVLPENVLQKAQSELLNWRGSGMSVME+SHRGKFTSIIDKAEADRLTLNI 128

Query: 63 PDNYKVIPLQGGASLOFSMIPLNATARGIKAY-YHVGSGWGERKSLYRGCK 110  
P + Y V+PLQGGAS QFS IPLN+ A Y V GSWG+K+ K

40 Sbjct: 129 PSDYTVLPLQGGASTOFSALPILNLCITPDSAVDYIVTGSNGDKAAKEAAK 177

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 711**

45 A DNA sequence (GBSx0755) was identified in *S. agalactiae* <SEQ ID 2191> which encodes the amino acid sequence <SEQ ID 2192>. Analysis of this protein sequence reveals the following:

Possible site: 24  
>>> Seems to have an uncleavable N-term signal seq

50 ----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 712

- 5 A DNA sequence (GBSx0756) was identified in *S.agalactiae* <SEQ ID 2193> which encodes the amino acid sequence <SEQ ID 2194>. This protein is predicted to be phosphoserine aminotransferase (serC). Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3380(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 10049> which encodes amino acid sequence <SEQ ID 10050> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- ```

20 >GP:AAF94318 GB:AB004196 phosphoserine aminotransferase [Vibrio cholerae]
    Identities = 104/210 (49%), Positives = 152/210 (71%), Gaps = 3/210 (1%)

Query: 4  NNTIEGTSLYDIPKINEVPVIADMSSNIIAVKYKVDOPAMIYAGAQNIGPAGVTVIIR 63
      N TI+G + D+P T++ P++ADMSS IL+ + V + +IYAGAQNIGPAG+ + I+R
Sbjct: 170 NETIDGIEINDLPVTDK-PIVADMSTILSREIDVSKYGVITYAGAQNIGPAGICIAIVR 228

30 Query: 64  EDMIN-ESEPTLSSMLDYKIQSDAGSLYNTPPAYSIIYIAKLVFHWKSLGQVDAMEKANRE 122
      +D+++ L +L+YKI ++ S++NTPP ++ Y++ LVP+W+K+ GGV A+E+ NR
Sbjct: 229 DULLDLASDLLAGVILNYKILARQSSMFNTPTTFAWYLSGLVFPQHLKAQGQGVKALIEVNRA 288

30 Query: 123  KSGLLYDYIDSSRFYSNPFVKKRSILCNIPFITINKDLDEKFKVEATERGPKNKGHR 182
      K+ LLY YIDSS+FY N + +RSL N+PF +LD+ P++ A RG ++KGHR
Sbjct: 289 KAALLYGYIDSSDFYRNEIH-PNRSILNVPFQLAKPELDOTFLELARAGLVSLKGHRV 347

35 Query: 183  VGGMRASLYNAFFPKQGVIELIDFMKTFEAR 212
      VGGMRAS+YNA P +GV L+DFMK FEA+
Sbjct: 348 VGGMRASLYNAPLEGVQALVDFMKKFEAQ 377

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 713

A DNA sequence (GBSx0757) was identified in *S.agalactiae* <SEQ ID 2195> which encodes the amino acid sequence <SEQ ID 2196>. Analysis of this protein sequence reveals the following:

- ```

45 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0466(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```



A related GBS nucleic acid sequence <SEQ ID 10047> which encodes amino acid sequence <SEQ ID 10048> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:CA87301 GB:ALL39079 putative acetyltransferase [Campylobacter  
jejunil]  
Identities = 46/170 (27%), Positives = 78/170 (45%), Gaps = 13/170 (7%)

Query: 7 IRLAFTNEIDQIMLIRARAZIAKTGSDQWQKEDGYPNRNDIIDDILNGYANVGIEDGM 66  
I+ A +++ I+ + ++A + OW ++ YPN +DI +V E+

10 Sbjct: 6 IQKAVNKDLASILEITKDALNMMKTMNPHQW--DENYTNIEIVPQEDIQAGRLVVFKEDE 63

Query: 67 LATYAVIDGHE-EVYDAIYEGKWHDNHRYLTPHRIALSNOFPRGEGLAQTFLQGL--- 121  
+ + + + + E Y + K D YL HR+A+ +G+G+AQ L

15 Sbjct: 64 ILGPTICNEKFKPEFYKQVIFNNKDYDDKAFYL--ERLAVKQNAKGKGVQKLLNFCENFA 121

Query: 122 IEGHKGEDFRCDTHEKNVIMQHILNKLGYSQYCKQVPLDGVR---LAYQKI 168  
+E HK R DTH KN N + KL + +CG + + LAY+KI

Sbjct: 122 LENHKA-SLRADTHSIGNFMMNSLFKKLDFNFCGNFDIPNYQDPPLAYEKI 170

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 714

A DNA sequence (GBSx0758) was identified in *S.agalactiae* <SEQ ID 2197> which encodes the amino acid sequence <SEQ ID 2198>. Analysis of this protein sequence reveals the following:

Possible site: 28  
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2968 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 715

A DNA sequence (GBSx0759) was identified in *S.agalactiae* <SEQ ID 2199> which encodes the amino acid sequence <SEQ ID 2200>. This protein is predicted to be D-3-phosphoglycerate dehydrogenase (scrA). Analysis of this protein sequence reveals the following:

Possible site: 54  
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3102 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10045> which encodes amino acid sequence <SEQ ID 10046> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

5 >GP:AA999020 GB:U67544 phosphoglycerate dehydrogenase (serA)
 [Methanococcus jannaschii]
 Identities = 102/313 (32%), Positives = 168/313 (53%), Gaps = 21/313 (6%)

Query: 31 ENPDYAIIRSONLNGQF---PSNLKAIARAGAGTGNIPITIEASAGIIVVNTFGANANA 87
++ D ++RS ++D LK I RAG G +NI +E A+ +GI+V N P A++ +
10 Sbjct: 40 KDAVVLVVRSGTKVTRDVEIKAEKLVKIGRAGVGVNDIDVEATEKGIIIVNAPASSIS 99

Query: 86 VIKRVIAALLLSARDYLGNRRVNNLTGIDIPKQIENGKAFAGNEIAGKKLGIVIGEGAI 147
V E + +L +AR N T K+ E +K F G E+ GK LGVIGIG I
15 Sbjct: 100 VAEITWGLMLAAAR-----NIPQATASLRGESEWDRKRFKGIELVKTGLVIGIGRI 150

Query: 148 GARIANDARRLGMIVLYDPTVYSIETAMNIISSHVQRVKEIKDIFETCDYITIHVPLINET 207
G ++ A+ GM ++GVDPY+ E A ++ V+ V I ++ + D+IT+HVPLT+ T
20 Sbjct: 151 GQVVRKAKAFGMNIIIGYDPTVYIPEVAESMG--VELVDDINELCKRADFITLHVPLTPKT 206

Query: 208 KHTPDAKAFSIIKKOTTINPARAEALVNNQSLFPAETIGVVKRYITDPGDKK-----LL 261
+H + ++MKK I+N AR L++ + L+EA++ G ++ D ++E LL
25 Sbjct: 209 RHIGREQIALMKQNAIIVNCARGGLIDEKALYEALKEGKIRAAALVFESEPPDKNPLL 268

Query: 262 NQRGITVPHVVGSTDEARLNCALMASQTIRCFMFTGTEITNSVMFNVHIOQTAPFR-IT 320
+ PH G ST+EA+ + ++ I + N VN PN+ Q + +
30 Sbjct: 269 TLINNVIGTPHQGASTEEAQKAAGTIVAEQIKKVLREGELAEVNVNMFNPQKELGKLKPYM 328

Query: 321 LNNKVVNIVAKI 333
L+ + + NIV ++
30 Sbjct: 329 LLAELVNVIMQV 341

```

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 35 Example 716

A DNA sequence (GBSx0760) was identified in *S.agalactiae* <SEQ ID 2201> which encodes the amino acid sequence <SEQ ID 2202>. This protein is predicted to be methylated-DNA--protein-cysteine S-methyltransferase (ogt). Analysis of this protein sequence reveals the following:

```

40 Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2460 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

50 >GP:AAF96913 GB:AB004427 methylated-DNA--protein-cysteine
 S-methyltransferase [Vibrio cholerae]
 Identities = 73/156 (46%), Positives = 99/156 (62%), Gaps = 9/156 (5%)

Query: 7 YOSPLGEIRLLADNLGLSGLYFVGOKYDMLAVNBEIVNMSNSTYLLGK--KWLDAYFSQ 64
Y SPLG + L A + GL G++F Q E + + +L K + LD YFS
55 Sbjct: 7 YSSPLGHWTLQASOGLGIVWEATO----TTQPEHLGDYVKECPILNKITRQLEDFYFSG 61

Query: 65 QNLP-SIPLSLRGTAPOTRVWORLQKIPFGDTKTYGELAKKL-NQOASAQVGAIGKNSI 122
Q +PL+ GTAPQ VW L KIP+G+ +Y +LA+ + N ++ +AVG A GEN I

```

-813-

Sbjct: 62 QRTQPELPLAASGTAPQSQVWHAICKIPTGRTWSYQQLAKRAIGNPKAVRAVLGANGNP1 121

Query: 123 SLIIPCHRVLRVGGQITGYAGGLERKSWLLRYEKK 158  
 S+I+PCHRW+G+ GQLTGYAGGLERK++LLR EK +

Sbjct: 122 SIIVCHRVVGGNGQLTGYAGGLERKAFLLLEKER 157

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 10 Example 717

A DNA sequence (GBSx0761) was identified in *S.agalactiae* <SEQ ID 2203> which encodes the amino acid sequence <SEQ ID 2204>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial cytoplasm | --- | Certainty=0.3137 (Affirmative) | < succ> |
| bacterial membrane  | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA07204 GB:AP001518 arsenate reductase [Bacillus halodurans]  
 Identities = 56/107 (52%), Positives = 74/107 (68%), Gaps = 1/107 (0%)

Query: 3 TFVEYPKCTTCSAKKELTELGLTFEADIKSNPPKVSLLKELLENSPYDLKKFPNTSGN 62  
 TFY+YPK TC+ AKK L + G+ ++ I PP LK+L E S +LKKFPNTSG

Sbjct: 4 TFYQYKCTGTCQKAKWLDQHGIEVNSVHIVEQPPSKBELKQLYEQSGLELKKFPNTSGK 63

Query: 63 SYRELGLKDKFDLITLDQLDILLASDGMILKRP/LKDKNKLQIGYR 109  
 YRELGLKDK + + D+ L+ L+SDGMILKRP+L +K+ +G++

Sbjct: 64 KYRELGLKDKVKEASEDELETLASDGMILKRPILITGDKV-TVGFK 109

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2205> which encodes the amino acid sequence <SEQ ID 2206>. Analysis of this protein sequence reveals the following:

Possible site: 38  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial cytoplasm | --- | Certainty=0.3969 (Affirmative) | < succ> |
| bacterial membrane  | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |

An alignment of the GAS and GBS proteins is shown below:

Identities = 64/99 (64%), Positives = 79/99 (79%)

Query: 19 ELTELGLTFEADIKSNPPKVSLLKELLENSPYDLKKFPNTSGNSYRELGLKDKFDLITL 78  
 EL +L FEADIK+NPVK LK +E S Y +K FNTSGNSYRELGLKDK D L+L

Sbjct: 3 ELKQLNSDFEADITKANFPKQDLKHAMKTSGYTINKFPNTSGNSYRELGLKDKIDQLSL 62

Query: 79 DQALLDLSADGMILKRP/LKDKNKLQIGYRKYKIDNL 117  
 D+A +L+L+DGMILKRP+L+KD +LQ+GYR Y++L+L

Sbjct: 63 DKARLLATDGMILKRPILIKDNGVLQVGYRKYKQELDL 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 718**

A DNA sequence (GBSx0762) was identified in *S.agalactiae* <SEQ ID 2207> which encodes the amino acid sequence <SEQ ID 2208>. This protein is predicted to be exodeoxyribonuclease (exoA). Analysis of this protein sequence reveals the following:

```

5 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.1859 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:AAA26879 GB:J04234 exodeoxyribonuclease [Streptococcus pneumoniae]
 Identities = 217/275 (78%), Positives = 245/275 (88%)

 Query: 1 MKLISWNIDSLNALTSBSTRAIMSRQVIDTLVARDADIATQETKLSAKGPTKQHLEVL 60
 MKLISWNIDSLNALTS+ S RA +S++V+ TLVAE+ADIIATQETKLSAKGPTKQH+E+L
 Sbjct: 1 MKLISWNIDSLNALTSBSARAKLSEVLQTLVAENADIATQETKLSAKGPTKQHVEIL 60

20 Query: 61 ETYPFSDYLWVRSSVSPARKGYAGTMFLYRKGLNPTVSFPEIDAPTMNDEGRITILE 120
 E FP Y+ WRSS SPARKGYAGTMFLY+K L P +SFPEI AP+IND EGRITILE +
 Sbjct: 61 EELFPYGYENTWRSSQEPARKGYAGTMFLYKCKELTPTISFPEIGAPSTMDLGRITILEFD 120

25 Query: 121 NCYITQVYTPNAGDGLKRLADRDQIMDIKYAEYLATLDSQKPLVATGDYNVAHKIDLANP 180
 ++TCVYTPNAGDGLKRL +RQ+ND KYAEYLA LD +KPLVATGDYNVAH EIDLANP
 Sbjct: 121 AFFVITQVYTPNAGDGLKLEERQVMDAKYAEYLAELDKKPLVATGDYNVAHNEIDLANP 180

30 Query: 181 SSNRRSAGFTAEERQGFITNLLAKGFTDTFRLHGDVFNIVYSWNAQRSTSKINNTGWRID 240
 +SNRRS GFT EER GFITNLLA GFITDTR++HGDVP Y+WWAQRS+TSKINNTGWRID
 Sbjct: 181 ASNRRSPGFTDSERAGFTNLLATGFTDTFRHVHGDVPERVYTWNAQRSTSKINNTGWRID 240

 Query: 241 YWLTSNRVADKITKSMIHSGDRQDHTPIILEIEL 275
 YWLTSNR+ADK+TKS+MI SG RODHTPI+LEI+L
35 Sbjct: 241 YWLTSNRADKYTKSMIDSGARQDHTPIVLEIDL 275

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2209> which encodes the amino acid sequence <SEQ ID 2210>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.2181 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

 Identities = 221/275 (80%), Positives = 251/275 (90%)

50 Query: 1 MKLISWNIDSLNALTSBSTRAIMSRQVIDTLVARDADIATQETKLSAKGPTKQHLEVL 60
 MKLISWNIDSLNALTS ES RAL+SR V+DTLVA+DADIIATQETKLSAKGPTKQH+E L
 Sbjct: 1 MKLISWNIDSLNALTSBSTRAIMSRQVIDTLVARDADIATQETKLSAKGPTKQHVEIL 60

 Query: 61 ETYPFSDYLWVRSSVSPARKGYAGTMFLYRKGLNPTVSFPEIDAPTMNDEGRITILE 120
 +YFP Y VWRSSVSPARKGYAGTMFLY+ LNP++FPPI APTMD EGRITILE E
55 Sbjct: 61 LSYFNYLHWVRSSVSPARKGYAGTMFLYKNTNPVITTFPKIGAPTMDEGRITILEFE 120

 Query: 121 NCYITQVYTPNAGDGLKRLADRDQIMDIKYAEYLATLDSQKPLVATGDYNVAHKIDLANP 180
 + ++TVYTPNAGDGL+RL DRQIMD KYA+YL LD+QKPLVATGDYNVAHNEIDLANP
60 Sbjct: 121 DFFVITQVYTPNAGDGLKRLDRQIMDKYADYITKLDAKPLVATGDYNVAHKIDLANP 180

```

-815-

Query: 181 SSNRSAGFTAERQGFTHLLAKGFTDTFKYLGEDVPNVYSWWAQRSRTSKINNTGWRID 240  
 +SNRRS GFT EERQGFTHLLA+GFTDTFR++HGD+P+VY+WWAQRS+TSKINNTGWRID  
 Sbjct: 181 NSNRRSPGFTDERQGFTHLLAKGFTDTFRHVHGDIPHVYTWAAQRSRTSKINNTGWRID 240

5 Query: 241 YWLTSNRFVADKITKSEMIHSGERQDHTPIILEIRL 275  
 YWL SNR+ DK+ +SSEI SG+RQDHTPI+L+L+L  
 Sbjct: 241 YWLASNRLVDKVRSEMISSGERQDHTPIILLDIDL 275

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 719

A DNA sequence (GBSx0763) was identified in *S.agalactiae* <SEQ ID 2211> which encodes the amino acid sequence <SEQ ID 2212>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.96 Transmembrane 28 - 44 ( 22 - 49)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4185 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8645> which encodes amino acid sequence <SEQ ID 8646> was also identified. Analysis of this protein sequence reveals the following:

Lipo: Possible site: -1 Crend: 5  
 MoG: Discrim Score: 17.78  
 GvH: Signal Score (-7.5): -4.56  
 Possible site: 55  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 1 value: -7.96 threshold: 0.0  
 INTEGRAL Likelihood = -7.96 Transmembrane 8 - 24 ( 2 - 29)  
 PERIPHERAL Likelihood = 9.28 138  
 modified ALOM score: 2.09

\*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.4185 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AD11512 GB:U60828 unknown [Lactococcus lactis]  
 Identities = 53/240 (22%), Positives = 102/240 (42%), Gaps = 24/240 (10%)

Query: 65 PTLIPGSSATQERFNSMLAQL---NQMGKIKSVLGLTVKKDMSLIYNGQIGSDHKPY 120  
 FTI I GS ++ +L N +K V+ + K+ + QGIS ++ P  
 Sbjct: 64 PTIYIGSGGNVTSIDWLVERLLPFIKNISSQKSLVMSNITRNKYLKVGQISQDKYPI 123

Query: 121 IVIGFENNEDGYSNIKKQTKWLQAMNDLQKKYKGRFNAIGHSNGLSWTIFLEDYDS 180  
 I G ++ + +K LQ + L + Y+ N +G+S+G ++ D ++  
 Sbjct: 124 IEFA---TVKGISGSLPSKGLQKLIIVLTENYQVFWINLVGYSSGATGAVVYMDTGN 180

Query: 181 DEFD-MKSLLTMTGTPNFNFEES-----NTSN-----HTQMLKDLINSGNIPSSIMVY 226  
 F + +++ +N E + + SN T+M + + N + S +  
 Sbjct: 181 PNFPVVKYISVLGDGEYNNETINLQGESLNVLKEGPIKVTENYQYIADNYQKVSSTQML 240

Query: 227 NLAGT--NSYDGDKIVFPASVETGKIVFOETAKHYTOLITVGNATHSDLEDNPEVQIV 284  
 L G + D +P+A + +F++ T T+ +HS P NP V++V  
 Sbjct: 241 LLEGFNSEKQTDGAIPLWADSPSIYHLFKKNGRETT-TLYPTKTSHQAPNPVTKYV 299

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8646 (GBS219) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 3; MW 31.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 7; MW 56kDa).

GBS219-GST was purified as shown in Figure 203, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 720

A DNA sequence (GBSx0764) was identified in *S.agalactiae* <SEQ ID 2213> which encodes the amino acid sequence <SEQ ID 2214>. This protein is predicted to be PTS system, cellobiose-specific IIC component. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -7.64 Transmembrane 263 - 279 (260 - 282)
 INTEGRAL Likelihood = -6.26 Transmembrane 200 - 216 (197 - 226)
 INTEGRAL Likelihood = -5.95 Transmembrane 157 - 173 (156 - 175)
 INTEGRAL Likelihood = -5.79 Transmembrane 307 - 323 (306 - 332)
 INTEGRAL Likelihood = -5.68 Transmembrane 131 - 147 (126 - 148)
20 INTEGRAL Likelihood = -4.73 Transmembrane 375 - 391 (370 - 396)
 INTEGRAL Likelihood = -3.61 Transmembrane 101 - 117 (98 - 119)
 INTEGRAL Likelihood = -1.75 Transmembrane 326 - 342 (324 - 342)
 INTEGRAL Likelihood = -0.37 Transmembrane 25 - 41 (25 - 41)
 INTEGRAL Likelihood = -0.16 Transmembrane 71 - 87 (71 - 88)
25

----- Final Results -----
 bacterial membrane --- Certainty=0.4057 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
30

The protein has homology with the following sequences in the GENPEPT database:

>GP:AACT4807 GB:AE000268 PEP-dependent phosphotransferase enzyme II
 for cellobiose, arbutin, and salicin [Escherichia coli K12]
 Identities = 60/197 (30%), Positives = 83/197 (41%), Gaps = 12/197 (6%)
35
Query: 209 LAIFLTLGSLFVFDIL--FRPYSYFSVSENLNAALSOHTDKIPYLYTFYTKNSPAMFG 266
 LA+ +G+ P L Y + V L A + H P L +SF G
Sbjct: 253 LALTALDNGIMTFVALENIATYCCQYGSVEAALAAAGKTFHIAKPMV-----DSPIFLG 305

40
Query: 267 GIGILLSLFLAVIYESRKLOSKNYKILATLITLITPLIDQNLFFVGLPVLITLIPILPW 326
 G G L L LA+ SR+ +Y + L L IF N P L GLP+I+ P++FIP V
Sbjct: 306 GSGATLGLLALFIASRA---DYRQVAKIALPSGIQINEPILFGLPIINMFVPIFV 362

Query: 327 LTTIFAEPGALMLYLKFVDPVAVTVTPSGTPLSLGFLASNGDWRYLPVTAILLVGGFTI 386
 L A Y+ + P P P+ L F +NG L V L + I
45
Sbjct: 363 LVQPIILALTLAAYMGILIPFVNTIAFWIMPTGLGAFFPNTNSVAALLVAFNLGIATLI 422

Query: 387 YRPFVKIAFAKEEQYEK 403
 Y PFV +A + +K
50
Sbjct: 423 YLPFVVVANKAQAIDK 439

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 721**

A DNA sequence (GBSx0765) was identified in *S.agalactiae* <SEQ ID 2217> which encodes the amino acid sequence <SEQ ID 2218>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1991(Affirmative) < succ>
10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 722**

A DNA sequence (GBSx0766) was identified in *S.agalactiae* <SEQ ID 2219> which encodes the amino acid sequence <SEQ ID 2220>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 39
 >>> Seems to have no N-terminal signal sequence

 INTEGRAL Likelihood = -5.79 Transmembrane 188 - 204 (179 - 206)
 INTEGRAL Likelihood = -5.36 Transmembrane 105 - 121 (104 - 127)
 INTEGRAL Likelihood = -4.41 Transmembrane 212 - 226 (210 - 229)
25 INTEGRAL Likelihood = -3.45 Transmembrane 72 - 88 (69 - 89)
 INTEGRAL Likelihood = -0.48 Transmembrane 124 - 140 (124 - 140)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8647> which encodes amino acid sequence <SEQ ID 8648> was also identified. Analysis of this protein sequence reveals the following:

```

35 Lipop Possible site: -1 Crend: 6
 SRCFLG: 0
 McG: Length of UR: 5
 Peak Value of UR: 2.99
 Net Charge of CR: 4
 McG: Discrim Score: 6.88
40 GvH: Signal Score (-7.5): -2.86
 Possible site: 30
 >>> Seems to have an uncleavable N-term signal seq
 Amino Acid Composition: calculated from 1
 ALOM program count: 5 value: -5.79 threshold: 0.0
45 INTEGRAL Likelihood = -5.79 Transmembrane 179 - 195 (170 - 197)
 INTEGRAL Likelihood = -5.36 Transmembrane 96 - 112 (95 - 118)
 INTEGRAL Likelihood = -4.41 Transmembrane 203 - 219 (201 - 220)
 INTEGRAL Likelihood = -3.45 Transmembrane 63 - 79 (60 - 80)
 PERIPHERAL Likelihood = 0.10 18
50 modified ALOM score: 1.66
 icml HYPID: 7 CFP: 0.331

 *** Reasoning Step: 3

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.3314(Affirmative) < succ>

```

-818-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2221> which encodes the amino acid sequence <SEQ ID 2222>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have a cleavable N-term signal seq.

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -11.20 | Transmembrane | 179 - 195 ( 173 - 201) |
| INTEGRAL | Likelihood = -3.66  | Transmembrane | 96 - 112 ( 95 - 113)   |
| INTEGRAL | Likelihood = -1.44  | Transmembrane | 203 - 219 ( 203 - 219) |
| INTEGRAL | Likelihood = -0.96  | Transmembrane | 115 - 131 ( 115 - 131) |
| INTEGRAL | Likelihood = -0.64  | Transmembrane | 63 - 79 ( 63 - 79)     |

----- Final Results -----

bacterial membrane --- Certainty=0.5479 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 160/228 (70%), Positives = 185/228 (80%)

Query: 10 MSKSHRQYQIYBGLRCVALCFISGYNAFTYVTQKRFAGVQTGNLLSFAIHLNKH Y 69  
 MSKK + YQ+YBGLRCA+ LCFISGY+NAFTY+TQKRFAGVQTGNLLSFAI LS +  
 Sbjct: 1 MSKKRKHRYQYBGLRCWTLCTFISGYNAFTYVTQKRFAGVQTGNLLSFAIRLSEQL 60

Query: 70 SQALAPLLPIMVFMGLQGSFTYPMKRNANKHQLHWYLLSSFALTQVAIVTIIITFLPSPF 129  
 +AL FLFP++VFMGLQGSFTYPMKRNANKHQLHWYLLSSFALTQVAIVTIIITFLPSPF  
 Sbjct: 61 KEALQPLFMIVFMGLQGSFTYPMKRNANKHQLHWYLLSSFALTQVAIVTIIITFLPSPF 120

Query: 130 TVAGLAFASIQVDTFKSLRGAPYANMMTGNIKQDAAYLLTKGLYEKNSDIFLIARNTII 189  
 TVA LAFASIQVDTFKSLRGAPYANMMTGNIKQDAAYLLTKGLYEKNSDIFLIARNTII  
 Sbjct: 121 TVAALAFASIQVDTFKSLRGAPYANMMTGNIKQDAAYLLTKGLYEKNSDIFLIARNTII 180

Query: 190 IIGGFIFGVVVCSTYPSKLGWNSLILIPLLYVILLGHEFYNLQVE 237  
 +I F GVVCST GE++L IL+FLLYVN LL EPY++Q +  
 Sbjct: 181 VILAFVGVVVCSTLILCIAGGYALMFLMPLLYVNYLLAQEFYHIQTQ 228

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 723

A DNA sequence (GBSx0767) was identified in *S.agalactiae* <SEQ ID 2223> which encodes the amino acid sequence <SEQ ID 2224>. This protein is known to be tellurite resistance protein. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -0.00 | Transmembrane | 190 - 206 ( 190 - 206) |
|----------|--------------------|---------------|------------------------|

----- Final Results -----

bacterial membrane --- Certainty=0.1001 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>



The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC22923 GB:U32807 tellurite resistance protein (teh)  
[Haemophilus influenzae Rd]  
Identities = 164/282 (58%), Positives = 205/282 (72%), Gaps = 1/282 (0%)

5 Query: 7 LLPTKMPVMTAQSIKAFLEKHNTKBTWAKLTLSGSLVPYQLSPGGEISRHIFDAS 66  
L+ YK NPWT+ ++F+ F EKHNTK GTW KLT+L G L PY+L+ +G+ I+ HIF  
Sbjct: 5 LICYKMPVMTKDNLPQMFGKHTKGTWGLTVLKGKLPYRLATENGDIABHIFTPE 64

10 Query: 67 SDIPYVDPQWHRVSPNSPDLSCYLTPYCKEDYFHKYGLTTRTHSRVIAAPMLSEKN 126  
S IPPV+PQ WH+V S DL C L PYC+KEDYF KKY T H +V+ +A +A  
Sbjct: 65 SHIPPVEPQWHRVZALSDDLECTLGPYCKKEDYFSSKYNNTAIHSDVDAAKIISP-CK 123

15 Query: 127 ILDLGCGQGRNSLYLSLGHQVTSVDSNGCSLVALENMKEELPYNIKRYDINTAIEG 186  
+LDLGGCGGRNSLYLSLIG+ VTS DN S+ L +E L + YDIN A I+  
Sbjct: 124 VLDLGGCGGRNSLYLSLGYDVTSMWHDNENSIAPINETKREKNTINISTALYDINANNIOE 183

20 Query: 187 HYDFILSTVFMFLNPDICSDILCMQSHTOIGGYNLIVSAMDTAENPCPLFPFTFKEG 246  
+YDFI+STVFMFLN + + II M+ HT +GGYNLIV+AM T +CPLFP FT F E  
Sbjct: 184 NYDFIVSTVFMFLNPRVPSIIKNMKEHTIVGGYNLIVAMSTDDVPCPLFPSTFAEN 243

25 Query: 247 QLKSTYDWEILIKYNNLGLHVRVDENGNRLKQFATLARK 288  
+LK YY DWE ++YNN+GLH+ DENGNR+K++FAT+LARK  
Sbjct: 244 ELKEYKDWEFLEYNNMGLHKTDEGNRHKQFATLARK 285

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2224 (GBS95) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 3; MW 35.6kDa) and in Figure 12 (lane 4; MW 35.6kDa). The GBS95-His fusion product was purified (Figure 191, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 292), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 724

A DNA sequence (GBSx0768) was identified in *S.agalactiae* <SEQ ID 2225> which encodes the amino acid sequence <SEQ ID 2226>. This protein is predicted to be methionyl-tRNA synthetase (metS). Analysis of this protein sequence reveals the following:

Possible site: 47  
>>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -0.32 Transmembrane 473 - 489 ( 473 - 489)

----- Final Results -----  
bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10043> which encodes amino acid sequence <SEQ ID 10044> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA811814 GB:Z99104 methionyl-tRNA synthetase [Bacillus subtilis]  
Identities = 395/667 (59%), Positives = 501/667 (74%), Gaps = 12/667 (1%)

50 Query: 20 EKSKFYITPIYPSGKLHIGSAVTTIACDVLARYKRMGFDVQYLGLDEHSGKIQCKA 79  
E +FYITPIYPSGKLHIG AVTT+A D +ARYKR+ GFDV+YL/G DEHSGKIQCKA  
Sbjct: 4 ENTFYITPIYPSGKLHIGAVTTVAGDAMARYKRLGFDVRYL/GTDEHSGKIQCKA 63

-820-

Query: 80 EEAGITPQEVVDGMASVKTWELLDISYDAFIRITDTYHREAVAKIFBQLLAQDDIYL 139  
 B+ ITPQEVVD A ++ LK+ L+IS D FIRIT+ H+ + K+P+LL GDIYL  
 5 Sbjct: 64 BQENITPQEVVDRAADDIQLWQLLEISNDQFIRITEKRHKVVIKVFQGLDNDIYLD 123

Query: 140 EYTGWYSVDEEPTFSQLAEVYREHNMIGGVAP-SCHEVKEVBEESYFFRMKKAOR 198  
 EY GNTS+ DE F+TE+GL ++ R+H G +IGG +P SCH VE + EESYFFRM KKAOR  
 10 Sbjct: 124 EYEGWYSIPDETFTTITQLVDIERNHKGEVIGSGSDSGHPVLIKESYFFRMKKAOR 183

Query: 199 LKAYIAHPEFIQDGRNEMLNKPIEGLEDLAVSKRTTITWGVQVPSNKHVITWIDA 258  
 L YY E+P FIQ+ R NEM+ NFI+QLEDLAVSKRTT WGV+VE NEKHVITWIDA  
 15 Sbjct: 184 LKCYTEENPTFIQDSSRNEMINNFIKQGLDLAVSKRTTFQGVKVFENFKHVITWIDA 243

Query: 259 LRFYIYALGYSGDGLSQYHKFWPADIHMIQKDLIRFHSIYWPIMLALDLPLPKRLVAH 318  
 L NY++ALGY +D Y K+WPAD+H+GK+I+RPH+IYWPIMLALDLPLPK++ AH  
 20 Sbjct: 244 LRFYIYALGYDTEHD-ELYQKTPADVHLVQKHIVRPHITIYWPIMLALDLPLPKQVFAN 302

Query: 319 GWFPMQDGHMSKSKGNVYFPEMLVERFGLDPLRYILMSLFPVSGDGTPTTPEYVGRINYE 378  
 GW +M+DGRMSKSKGNV P L+ER+GLD LRYIL+R P GSDG PTTE +V RINI+  
 25 Sbjct: 303 GWMKMDGHMSKSKGNVDFVTLIERYGLDELRYILLREVFPVSGDVPTTPEGFERINVD 362

Query: 379 LANDGLNLNKTAMWIKYFDGEVPRF-AVAITDFADLASVATDSIENYHQMKAQVDFR 437  
 LANDGLNLNKT+AM+HKYFDG++ + T+FD L SVA ++++ Y K ME +F  
 30 Sbjct: 363 LANDGLNLNKTAMWIKYFDGQISYKGAITEFDHILTSVAESTVYKAKEMEMEFVS 422

Query: 438 ALAEVAVGLISRTNKYIDSTAPWVLANDETDRKLAAMVSHLVASLAVUHLIQPFMMETS 497  
 AL +M LISRNTNKYIDSTAPWVLAND ++L +VM HL SLR+ A L+QFP+ +T  
 35 Sbjct: 423 ALST+MLISRTNKYIDSTAPWVLANDPAKKEELRSMYHLASLRISAVLLQPLTKTP 482

Query: 498 DAINEQGLGL--GATFDLSKIT-FAQLPEGVYVVAKSGPIFPLRDMDEITTYIKEMVMAG 554  
 ++ BQLG+ + ++T F L + V KG P+PRL+ E+EI YIK +M G  
 40 Sbjct: 483 EKMFEQGLITDESLEKAWDSITAPQLKD--TKVQGEPLPRLAEAREIAYIKGMQ-GS 539

Query: 555 APVEKEVPEVEVETLSSQKIQKPEDFDAVEIKRVARVIEVEKVGSGSKILFRPLDAGEKH 614  
 AP ++E EE + +I + F VE+RVARVIE E V+ +D+L+ +LD G E  
 45 Sbjct: 540 APAKESTKEEPEVDRIPEITIDQFMVYKRVARVIRAEPPVKADRIQLDLGFE-K 598

Query: 615 RQLLSGIKAPYPMQGLVKKIKIVANLKPRMCKYVSCGMILSAEHGSKITVITVDSA 674  
 RQ++SGIAK Y E ELVQKIL V NLKP K ++ +SCGMIL+ E DG L V++D +  
 50 Sbjct: 599 RQVVSIGIAKHYTPE-ELVGKILCVTHLKPVK-LRGLSCGMILAGEADGVKVSIDGS 656

Query: 675 VANGSII 681  
 + G+ I  
 55 Sbjct: 657 LPKGTIRI 663

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2227> which encodes the amino acid sequence <SEQ ID 2228>. Analysis of this protein sequence reveals the following:

Possible site: 29  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1245 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 516/665 (77%), Positives = 573/665 (85%), Gaps = 4/665 (0%)

Query: 21 KKSFYITPIYYPGSKLHIGSAYTTIACDVLARYKRMGGFDVOYLAGDHEHQKIQKAE 80  
 KK FYITPIYYPGSKLHIGSAYTTIACDVLARYK+MG +V YLAGDHEHQKIQ KA+  
 60 Sbjct: 3 KKPFIYITPIYYPGSKLHIGSAYTTIACDVLARYKRLMGHEVFFYLAGDHEHQKIQKRAK 62

Query: 81 EAGITPQEVVDGMASVKTWELLDISYDFIRITDTYHREAVAKIFBQLLAQDDIYLGE 140  
 EAGITPQ YVD MA+ VK LN+LLDISYD FIRITD YHRE +V+FE+LLAQ DIYLYGE  
 65 Sbjct: 63 EAGITPQYVDNMAQDKVALQGLLDISYDTFIRITDTYHREAVAAVFEKLLAQDDIYLGE 122

Query: 141 YTCWYSVSDKEFPFESQLASVYRDENGNMIGGVAPSGHVEKVSERSYFFRMSKYADRLK 200  
 Y+GWYSVSDKEFPFESQL EV+RDE+G +IGG+APSGHVE VSESYF R+SKY DRL  
 5 Sbjet: 123 YSGWYSVSDKEFPFESQLKSVFRDHDGQVIGGIAPSGHVEWVSERSYFLRLSKYDRLV 182

Query: 201 AYYAEHPEFIQPDGRMNEM+KNPIEPGLKILAVSETTYTWGVPSGNPKHVIYVIMDALM 260  
 A++ E P+FIQPDGRMNEM+KNPIEPGLKILAVSETT+TWGV VPS+PKHV+YVIMDAL+  
 10 Sbjet: 183 APPKERPDFIQPDGRMNEMVKNPIEPGLKILAVSETTFTWGVVPSDPKHVYVIMDAL 242

Query: 261 NYISALGYGMSDDLQYKFKWPADI-HMIGKDILRPHSIYWPIMLMALDLPKRIVAHG 319  
 NY +ALGY ++ + + KFW + HM+GKDILRPHSIYWPI+IM LDP+P RI+AHG  
 10 Sbjet: 243 NYATALGYRQANH-ANFDKFWNGTVPFHMVGKIDILRPHSIYWPILIAMLDLPMPRLIAHG 301

Query: 320 WFWNQDGRMSKSGKNVYVPEMLVERFGDPLRYTLMRSLPVGSDGTFTPEDYVGRINYL 379  
 WFWN+DGRMSKSGKNVYVPEMLVERFGDPLRYTLMRSLPVGSDGTFTPEDYVGRINYL  
 15 Sbjet: 302 WFWNQDGRMSKSGKNVYVPEMLVERFGDPLRYTLMRSLPVGSDGTFTPEDYVGRINYL 361

Query: 380 ANDLGNLLNRTIAMVINKYFDGSEVPPRA-VATDFADLASVATDSIENYHKQMEAVDFPRA 438  
 ANDLGNLLNRT+AM+NKYFDG VP + T FADL+ + + +YHK MEAVD+PRA  
 20 Sbjet: 362 ANDLGNLLNRTVAMINKYFDGTVEAYVDMGTAFADLSQLINDQADYHGHMEAVDYPRA 421

Query: 439 LEAVNLIISRTNKYIDETAPWLAKDETDRDKLAAMVSHLVAIRVVAHLIQPFMMETSD 498  
 LEAVN +I+RTNKYIDETAPWLAK++ D+ +IA+VM+HL ASLR+VAH+IQPFMMETSD  
 25 Sbjet: 422 LEAVNTLIARTNKYIDETAPWLAKEDGKAQASVGHAAASLRVAHVHIQPFMMETSA 481

Query: 499 AIMEQLGLGATPDLKLTFAADLPBGVVRVARGSPIFRRLMEDEITYTKQMGAKA-PV 557  
 AIM QLGL DL L AD P +VARG+PIFRRLME EI YIK QM A  
 30 Sbjet: 482 AIMEQLGLGATPDLKLTFAADLPBGVVRVARGSPIFRRLMEDEITYTKQMGAKA-PV 541

Query: 558 EKEWVPEEVELTSSEKQIKFEDTFAVEIRVAEVEIEVEKVEGSDKLLRFRDLADGDESHRI 617  
 EKEWVPEV L S K I F E FFAVEIRVAE EV IVEGS+KLLRFR+DAGD RQI  
 30 Sbjet: 542 EKEWVPEEVALSKSEKQITFTFAVEIRVAEVEIEVEKVEGSEKLLRFRVADGDEGQDRQI 601

Query: 618 LSGIAKFPYNEQELVGKGLQIVANLKPRKMMKKYVSGMILSAEHGDKLTVLTVDSAVAN 677  
 LSGIAKFPYNEQELVGKGLQIVANLKPRKMMKKY+SGMILSAEH +LTVLTVDS+V N  
 35 Sbjet: 602 LSGIAKFPYNEQELVGKGLQIVANLKPRKMMKKYVSGMILSAEHGDKLTVLTVDSVVPN 661

Query: 678 GSIIG 682  
 GSIIG  
 40 Sbjet: 662 GSIIG 666

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 725

45 A DNA sequence (GBSx0769) was identified in *S.agalactiae* <SEQ ID 2229> which encodes the amino acid sequence <SEQ ID 2230>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2633 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-822-

**Example 726**

A DNA sequence (GBSx0770) was identified in *S. agalactiae* <SEQ ID 2231> which encodes the amino acid sequence <SEQ ID 2232>. This protein is predicted to be branched chain amino acid transport system II carrier protein (brnC). Analysis of this protein sequence reveals the following:

```

5 Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-14.91 Transmembrane 279 - 295 (269 - 303)
 INTEGRAL Likelihood = -9.98 Transmembrane 82 - 98 (74 - 102)
10 INTEGRAL Likelihood = -6.58 Transmembrane 345 - 361 (340 - 364)
 INTEGRAL Likelihood = -6.00 Transmembrane 157 - 173 (153 - 179)
 INTEGRAL Likelihood = -4.30 Transmembrane 48 - 64 (45 - 66)
 INTEGRAL Likelihood = -4.14 Transmembrane 251 - 267 (250 - 278)
 INTEGRAL Likelihood = -4.09 Transmembrane 308 - 324 (305 - 326)
15 INTEGRAL Likelihood = -2.55 Transmembrane 218 - 234 (216 - 237)
 INTEGRAL Likelihood = -1.38 Transmembrane 126 - 142 (126 - 142)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9407> which encodes amino acid sequence <SEQ ID 9408> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

25 >GP:AAC00400 GB:AF008220 branch-chain amino acid transporter
 [Bacillus subtilis]
 Identities = 130/367 (35%), Positives = 204/367 (55%), Gaps = 12/367 (3%)

30 Query: 1 MSSEKPSPSNFIATVILVLLTIGPLFAIPRTATVSPFZIGVAPIVGHSP--IALLCFPIACFF 58
 +++K P F F V+LHL+IGPLFAIPRT TVS+BIG P + P ++LL FT FF
 Sbjct: 73 LADTANFPVFGTIPTVVLHLSIGPLFAIPRTGTVSYETGAVPFLTGVPERRLLIIFTIFF 132

 Query: 59 AARYTLAIRPWGILDGVKILTPVFAFLISLWVVGAIAYGNLESANASADYAGKAPGSG 118
 YYLA+ P+ ++D VGNILTP+ F I+ ++V+ AI + Y G G
35 Sbjct: 133 GVTYYLALNPISKVDRVGNKILTPI-KFTI ILIIVLKAIFTPWGLGAVTEAYKGTVPVFKG 191

 Query: 119 VLAGVNTLLAALAAVAFCLVATETLAKPGPKTKKEYLSTIHTIVGIVTSLAPSILYIGLGL 178
 L GY T+DLA++ F +V +K G K + G++ +L +Y+ L +L
40 Sbjct: 192 FLBGTKTMDELASIVFGVVVNAVSKGVTSKGLAALAACTKAGVIAALGLTIPIYVSLAYL 251

 Query: 179 GNKFPVRADILADPNVNGKAVLSQASYKLGNGPGRYFISIMVTLTCFPTTTVGLIVSVSE 238
 G A V +GA +LS +S+ LFG+ G L +T+ C TT+GL+ S +
 Sbjct: 252 G-----ATSTNAIGPVGGAKILSASSHYLPGSLGNIVLGAATVCTTYSIGLVISCCQ 306

45 Query: 239 PFDKRNPRGNYLKFAIVFTLIGFLIANLGLNAVITPSVPLTLLYPIVIVITVGLILINKW 298
 + F K +YK+ T+ TL +IAN GL +I FSPV+L+ +Y+ IVI++ I+K
 Sbjct: 307 YPSKILFALSYKIVVITVITVLSLIINPGLAQIIAPSVPILSAIYPLAIVITVLSFIDKI 366

 Query: 299 LFLSKK--GNSLTIGLVTLVSVFVEVLGAGNQEKTLQVWGFLPFHTTSMGWLVPMLIGI 355
 ++ + GL +++ ++ NG L LP +++ G++P++G
50 Sbjct: 367 FKRRREVYIACILGVLPSILDGKA-AGPSLGLSLVPLNANLPLYSIGIGWLGLGVGA 425

 Query: 356 VFSLVLS 362
 V VLA+
55 Sbjct: 426 VIGYVLT 432

```

There is also homology to SEQ ID 2234.

A related GBS gene <SEQ ID 8649> and protein <SEQ ID 8650> were also identified. Analysis of this protein sequence reveals the following:

-823-

[illegible]

-824-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 727

20 A DNA sequence (GBSx0771) was identified in *S.agalactiae* <SEQ ID 2235> which encodes the amino acid sequence <SEQ ID 2236>. Analysis of this protein sequence reveals the following:

Possible site: 36  
>>> Seems to have no N-terminal signal sequence

```

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3291(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 A related GBS nucleic acid sequence <SEQ ID 10041> which encodes amino acid sequence <SEQ ID 10042> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
35 vaccines or diagnostics.

### Example 728

A DNA sequence (GBSx0772) was identified in *S.galactiae* <SEQ ID 2237> which encodes the amino acid sequence <SEQ ID 2238>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
40 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -8.33 Transmembrane 117 - 133 (112 - 136)
 INTEGRAL Likelihood = -3.77 Transmembrane 53 - 69 (53 - 70)
 INTEGRAL Likelihood = -3.40 Transmembrane 98 - 114 (97 - 115)

```

```

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.4333(Affirmative) < success
 bacterial outside --- Certainty=0.0000(Not Clear) < success
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

50 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 729

A DNA sequence (GBSx0773) was identified in *S.agalactiae* <SEQ ID 2239> which encodes the amino acid sequence <SEQ ID 2240>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.19 Transmembrane 22 - 38 (20 - 44)
----- Final Results -----
 bacterial membrane --- Certainty=0.2678 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8651> which encodes amino acid sequence <SEQ ID 8652> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1 Crend: 3
SRCFLG: 0
McG: Length of UR: 21
 Peak Value of UR: 3.11
 Net Charge of CR: 2
McG: Discrim Score: 11.30
GVH: Signal Score (-7.5): -5.35
 Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program count: 1 value: -4.19 threshold: 0.0
 INTEGRAL Likelihood = -4.19 Transmembrane 5 - 21 (3 - 27)
 PERIPHERAL Likelihood = 6.74 53
modified ALOM score: 1.34
icml HYPID: 7 CFP: 0.268

*** Reasoning Step: 3
----- Final Results -----
 bacterial membrane --- Certainty=0.2678 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15623 GB:Z99122 spore coat protein (inner) [Bacillus subtilis]
Identities = 71/359 (19%), Positives = 148/359 (40%), Gaps = 49/359 (13%)

Query: 127 ISYRGNTSRYPDKKSLKVKPFVINKLKKKKHRLAQNPKESSWVLGHGPFLLDRTLLRNYLSYN 186
 I+YRG+ R F KKS + F K + L+ + D +L+RN LS +
Sbjct: 47 IAYRGSHIRDPKKKSYHISFYQPKTPRGAREIH-----INASYKDPISMRNKLSD 97

Query: 187 IAGRIMSYPAPNVRYKELFVNGHYQGVYLAVENTDQGEQRVPTRKSKKHLRTPTTVAVMR 246
 R+ + +P + + +NG+ +GVYL +B+++ + +KL A D
Sbjct: 98 FFSSELGTLSPKARPAFVKMGKNGBGVYLELSVDH-----YYLAKRKLAJGAIFYAVDH 151

Query: 247 EHKAKQKLDNIVVHYHTEQSGISALDVKYFGKQRLTSKQLRFINKD----INIRKVLVSYD 302
 + D + ++L++ Y +++ +++ +F +D IN + K +
Sbjct: 152 DANFSLMSD-----LERTFTSLSELGY--EKKTGTGREDDFYLQDMIFKINTVPAKQF-- 202

Query: 303 FSQYPKYIDRESFANYFVINEFRPNVDAGKFSYLYKDLRDRA- KLVVDVFNNAFNQIE 361
 S+ K++D + + + F N D + LY+ +++ ND++ + I
Sbjct: 203 -SKVTKHVDVDKYLRLAGIVFTSNYDGFVNVALYRSGETGLFEVIMDIDATYWGRIH 261

Query: 362 GRVDEADPTLTDAPWFNMLTKDAFIDLVRHYRKLKGVLATETLSNYIDTRHPLGSA 421
```

-826-

G AD+ FN L YK L + L + + Y++ P  
 Sbjct: 262 GERMAADYVRIQQ--FNTLTAIILDESEFRKSYKLLLEKTLQSLFTIEYME-----PK 312  
 Query: 422 IDAFYKGGYVFDLKNMTPRNYLLPTEPN-VTSYHKSVEQLKDETKKRGWMDRNETL 479  
 I Y++ P + P +N + + + + +IK R + + + L  
 Sbjct: 313 IMMYER-----IRPFVIMDPYKKNDIERPDEPDI CEYIKNRSQYKHLISL 362

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 730

A DNA sequence (GBSx0774) was identified in *S.agalactiae* <SEQ ID 2241> which encodes the amino acid sequence <SEQ ID 2242>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 731

A DNA sequence (GBSx0775) was identified in *S.agalactiae* <SEQ ID 2243> which encodes the amino acid sequence <SEQ ID 2244>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -4.62 Transmembrane 5 - 21 ( 3 - 24)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2848 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA05949 GB:AP001514 unknown [Bacillus halodurans]  
 Identities = 199/697 (28%), Positives = 322/697 (45%), Gaps = 58/697 (8%)  
 Query: 57 KPFVVGVDVRESSLAGYHENDFPITQKTYREWFHLISNGANTVRVKVPMGVAFYDALYH 116  
 K + GV++ G + I +K Y WF I MG N +RV FY AL  
 Sbjct: 414 KKLQIHGVNLGAGKRGPTTFGEALIKEDYRWFQIGMGONAIRVYTLHPGPFYALKR 473  
 Query: 117 HNGASKRPLNLLQIRIDSYRNNASITAPNDYRNGYLKREAGVVDILIGRKQVNTDLG 176  
 +N+ + P+YL G+ ID ++ AP++ ++E K +VD++HG V + + G  
 Sbjct: 474 YNEQHENPIYLPFGVWIDREPLEDITLDAFDGETNEEPQQMKRIVDVHGHAVV-DPNEG 552  
 Query: 177 SRH--YHYDLSPVVLGYVGGDWNNGTVAYTNHQQK--QYGRGYFKTSVAANPPEVWLA 233  
 H Y D+SP+ G+++G +W TV TN Y G+Y +T A PPE LA  
 Sbjct: 533 HARGVYQADVSPYTTIGWILGIEWYPTHVKATKNKPNFDIGYDGRKYVETK-DABPPEYWLA 591



-827-

Query: 234 QVDELTHYETAKYQWHLISFSPSTTDPF-HYRKPFEQAQPKTVLAVENIQANSVK 292  
 D L YE +Y W +SF+N TTD H +P+E + V NV +++ + +  
 Sbjct: 592 NQDFLLSYLBIQYNIIRPVSYTNWVYTDLLTHPAPFWEDELDVGVDPNVIHLKGPA-TE 650

5 Query: 293 AGMFAYKATDFEPRYKDYLLFDGNISKEDRQKIKLSLSQVYKLLNAYHKIPVLVTG 352  
 FA+Y +P Y D+L ++++ I D + EL+ GY+K L+ H +P+L+  
 Sbjct: 651 TNQFASYHV---YPPYFDINRYEDYIHYVDR--GELANNYAGYLKDLHAHDLPILLAE 705

10 Query: 353 YGYSTANGIA-QKEIDKRFLPINEKQGGQLLEDUYESFISGSGGATTINAWQDDNARAW 411  
 +G +RG+ + K ++E+EQG+ ++E +E I G I WQD+W R W  
 Sbjct: 706 FGVPASRGUTHENPPGKQGLSSEEGKIVLFLFEDIERKILGGLIPTWQDEWFKTW 765

Query: 412 NTSFATNCHISQFLWDAQVFNQGGILLGFKNAKIHYYVDGKRGK-----EMKHPMTSA 466  
 NT N + W +AQ Q +GLL F K D + + E HP +  
 Sbjct: 766 NTMDYDNPDRPPFWSNAQTNRQQGGLSFDRLKVKVNGDQDWDASLLYEEDHPVTKR- 824

15 Query: 467 TQCDLYASDESSELYLAIKTKPELKE-----KRLPIDITPKSGSRHNGSK-VTFSEK 520  
 LY DE YLY I K + +L+D P G+ + + VTF  
 Sbjct: 825 ----LYMDHDERLYLFRIDMSGSTDDFFKDGFFILVLDTLPQGNHKEVEGVTPEH 880

20 Query: 521 SDFVLSIDPAKSELFVQERYNALKNVLAQLNGKDFYAPPPKQISSNPFQINVMRLNTK 580  
 DF+++ +S + V Y+ Y + + + P+M+ F++I+ L N+  
 Sbjct: 861 IDFIIEKGYDESRVKVDAYDPPFYQYSQIYQMIEETSIEPQNTVGVPKIHVAL-NQE 939

25 Query: 561 IVEDMEKVATERFLP--THPTGLLKTGTDTRHCKTFDSDQTD--ISPGKDFIEVRLPNQL 636  
 I ++ +T +P + TG L+ G D -DS D + K IEVRLPNL  
 Sbjct: 940 I-----RIPSTNEVIPPSSYYETCELRHKGQDPEADDYDGLADFFVNEKGMIEVRIWML 994

30 Query: 637 LNPSDFPSQKHDDYFKHYGVKELE-IESI-ALGLGANSKENTILKMD----- 683  
 L+F DPS +++ ++ G + E IE + A L K++ ++ D  
 Sbjct: 995 LSFKDPGQREVMSAIEBGCETSEIEIGVRAAVLPVEPKDDSYQVVDALPALDGLRL 1054

Query: 684 -----YRLQWNERPDTKTFLKDSYYSIKKEWSEKERE 714  
 Y + M+ P + LK SY +K+ ++ +E  
 Sbjct: 1055 DEVMNMYTWTMDIPLYEERLKQSYDLVKCAFTSIKE 1091

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8653> and protein <SEQ ID 8654> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 2  
 Mcg: Discrim Score: 12.00  
 Gvfi: Signal Score (-7.5): -5.46  
 Possible site: 21

45 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 1 value: -4.62 threshold: 0.0  
 INTEGRAL Likelihood = -4.62 Transmembrane 5 - 21 ( 3 - 24)  
 PERIPHERAL Likelihood = 7.32 223  
 modified ALOM score: 1.42

50 \*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.2848 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 55 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 2244 (GBS62) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 7; MW 80.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 4; MW 105kDa).

60 The GBS62-GST fusion product was purified (Figure 100A; see also Figure 193, lane 7) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure

100B), FACS (Figure 100C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 732

A DNA sequence (GBSx0778) was identified in *S. agalactiae* <SEQ ID 2245> which encodes the amino acid sequence <SEQ ID 2246> in others. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>> Seems to have no N-terminal signal sequence
10 INTEGRAL Likelihood = -7.48 Transmembrane 310 - 326 (302 - 335)
 INTEGRAL Likelihood = -7.32 Transmembrane 362 - 378 (361 - 380)
 INTEGRAL Likelihood = -7.11 Transmembrane 334 - 350 (329 - 355)
 INTEGRAL Likelihood = -2.28 Transmembrane 381 - 397 (380 - 397)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.3994 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

20 A related GBS nucleic acid sequence <SEQ ID 10039> which encodes amino acid sequence <SEQ ID 10040> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAB05950 GB:AP001514 unknown conserved protein in others
[Bacillus halodurans]
25 Identities = 143/405 (35%), Positives = 226/405 (55%), Gaps = 5/405 (1%)

Query: 11 IVPAYNESTTIVSSIDSLHLJDEYAEIIVVDGSSDNTSDVLKEFALMKISNTIDSII 70
+VPAYNE T I + ++ SLL L Y EI+VV+DG+D T +V+ E F ++K+ I I
30 Sbjct: 69 LVPAYNEETGLIETVRSLLSLKYQTETIVVVDGSDTQTLLEVIIEHFQVVKVKIRKQI 128

Query: 71 ATCTCKDVFORQVQVKVLTIVKENGKGDALMGINRANNDYFLCLDADSMQLVDLSQ 130
T+ K V+Q + L L+ K NGK DALH G+N + Y YF +D DS+L+ D+L + E
35 Sbjct: 129 ETEPIKGVYQSTIFP-HLLLVDSKNGKADALMGAGLVNPKYPFCSIDGSIETDALK 187

Query: 131 ISKSIQV----DPTVIAVGVLQVQAQGVKIEQKVASYRLPWRIIPCAQALEYDSSFLGA 186
+ K I + VIA GG V++A G I+ G V S +L + Q +EY +FL
35 Sbjct: 188 VMKPIVTSRDDSDRVTASGGNVRVANGSDICMSVLSVQLAKNPLVMQVIREYLRFLMG 247

Query: 187 RIFLDYLRANLLIISGAFGLFKKDLKAVGGYDTQTLGDMELVMKHLFFCRNNINPYRC 246
RI L LIISGAF +F K V GGY +T+GDMELV++LH + + RI
40 Sbjct: 248 RIGLSRHNMLIISGAFSVPFAKQNMVAGGYSKTIVGDMELVLRHLRVKRLKRRIT 307

Query: 247 YETDAVCWSQAPTNLGDRLKRRRWVGLGYCLKKYKSIPIANRFGAVGSISYIYILFE 306
+ D VCM++AP L++CR RW+ GL+ L ++ + N++G VG+ S Y+ + E
45 Sbjct: 308 FVPDPCVCTEAPATFRLVQRQSRNWHGLMESLNLHRCMTFNPKYGLVGTASIPYFVIVE 367

Query: 307 LLLPFIIECPGIVIIIFLSLLPNQINPPFISVLSVLYFYCYVLTITLSSFLHRTYSQQVIGI 366
P +E G + I + F L + F ++L L++ Y + ++++ + +S + +
50 Sbjct: 368 FPGFVVEIMGYLYIVPAFFPGLYVEPALALPLFLVLYGTVPSTAVILEGWSLKKRPKV 427

Query: 367 LDIVKVFYIAVFRYLILHLPVLTVPKVASVIGYKKNKMMVGHITRE 411
D+ ++ ++F L P+ + ++I + WG +TR+
Sbjct: 428 SDMSRLMIFSLFEALMYRPLTIVLRFGATIEALPKSKWGEIMTRK 472

```

55 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2247> which encodes the amino acid sequence <SEQ ID 2248>. Analysis of this protein sequence reveals the following:

Possible site: 60

-829-

```

>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -11.04 Transmembrane 33 - 49 (24 - 57)
INTEGRAL Likelihood = -10.77 Transmembrane 376 - 392 (370 - 399)
INTEGRAL Likelihood = -7.86 Transmembrane 344 - 360 (342 - 372)
5 INTEGRAL Likelihood = -4.94 Transmembrane 63 - 79 (55 - 81)
INTEGRAL Likelihood = -2.07 Transmembrane 403 - 419 (403 - 419)

----- Final Results -----
bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
10 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 64/397 (21%), Positives = 173/397 (43%), Gaps = 71/397 (17%)

15 Query: 6 FRKRSIVPAYNBSST-TIVSSSIDSLHLDYRAYRIIVDDGSSDNTSVLKEEFALMKISN 64
++ ++P+YNE +++ ++ S+L Y EI +VDDGSS+ + L KE+ ++
Sbjct: 90 YKVAAVIPSYNEDABSLLETLSGLVAQTYPLSEIYIVDDGSSNTDAQLIEBY---VHR 145

20 Query: 65 TIDSIATYTCCKVDFQRQVGKVLTLVKEGGKSDALMGINAAVDYFLCLDADSNLQ 124
+D C++V V +L+ N GK A ++ D FL +D+D+ +
Sbjct: 146 EVD-----ICRWI-----VHRSLV---NKGKHAQANAFERSADVFLTVDSDTIY 190

25 Query: 125 VDSLSQISQSIQVDPFTVIWGLGVQVQGVKIQGKVASRYLFWRIIPCQALEYDSSFL 184
+L+ ++ K+ D TV A G + + ++ + YD+P
Sbjct: 191 FVALBELKSPN-DETVAA-----TGLNARNRQNLN/RLTDIRYDAAF- 235

30 Query: 185 GARIFLDYLRANLII-SGAPGLFKD-LVKAVGVDYDTQT-----LGEOMELVMKLF 235
G L N++ S+ +++++ ++ + Y QT +G+D L
Sbjct: 236 GVERAAQSITQNLVCSGLSIYRREVIIPNLERYKNQTFGLFVSGIDDRCLT----- 289

35 Query: 236 CRSNIPY-RICYETDAVMSCAPTNLGLKQRPRVILGLY-CQLKIKYSIFAYYRFGA 293
N I R Y++ A C + P L KQ+ RN ++ + K I +N
Sbjct: 290 --NYAIDLGRTVQSTARCDTVFPQLKSYLKQCNWNSFFPKESIISVKILSN----P 343

40 Query: 294 VGSISYIYIYIFELLTPFIECFGIYIIFLSLLFNGLNIPFFISLVSLNIFYCV--LITS 351
++ + I+ ++ ++ ++ +LLFNQ +L+ L+ F + ++ L
Sbjct: 344 IVALNTIFEVVMFM-----LIVAIGNLLFNQ---AIQLDLIKLFAFLSIIFVALC 392

40 Query: 352 SFLLHRYSCQLVIGILDIVKVFYIAVRYLILHPLVT 388
+H + + + + ++ V + L L+ + T
Sbjct: 393 RNVHYMIKHPSFLLEPLVGIILHFLVLQPLKLYSLCT 429

```

A related GBS gene <SEQ ID 8655> and protein <SEQ ID 8656> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: -5.18
GvH: Signal Score (-7.5): -4.91
> Possible site: 14

50 >>> Seems to have no N-terminal signal sequence
ALOM program count: 4 value: -7.48 threshold: 0.0
INTEGRAL Likelihood = -7.48 Transmembrane 310 - 326 (302 - 335)
INTEGRAL Likelihood = -7.32 Transmembrane 362 - 378 (361 - 380)
INTEGRAL Likelihood = -7.11 Transmembrane 334 - 350 (329 - 355)
55 INTEGRAL Likelihood = -2.28 Transmembrane 381 - 397 (380 - 397)
PERIPHERAL Likelihood = 1.22 140
modified ALOM score: 2.00

*** Reasoning Step: 3

60 ----- Final Results -----
bacterial membrane --- Certainty=0.3994(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```



-831-

Query: 5 IRRARLGDEVNLAYIQTESWKAAPGKILPEDIIOKTEIEPAITMYQQLLHKVEGKGYIL 64  
 +R L D ++ I+ W++A+ ++D+ + A G+ ++  
 Sbjct: 10 VREMTLADCDRVSLIRVQWQSAIRGLMPYLDAMDPAADAERRSLFARPPGGRVNLV 69

5 Query: 65 EVDNPHCMAMD----KSRSDGMLDYAEILCHSLKEGMRGYGSGQNNHVLSEIQAG 120  
 D + W + E D AEL ++ +G G G + + + AG  
 Sbjct: 70 AEDEGGSEVVGWACHGPPYRIGRATAD-AELYALVDAARFGAGIGRATAGESVRRCRANG 128

10 Query: 121 YNKVILWVPENTNRRKFPYDRPGSPFKGSKTY 153  
 A +++LWV N RAR+PYDR GP G + +  
 Sbjct: 129 HARMLLWVLKGNVRRRFPYDRAGFRPDGAKDPY 161

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 734

A DNA sequence (GBSx0780) was identified in *S.agalactiae* <SEQ ID 2251> which encodes the amino acid sequence <SEQ ID 2252>. This protein is predicted to be a DNA-binding protein. Analysis of this protein sequence reveals the following:

20 Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 25 bacterial cytoplasm --- Certainty=0.1162 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 735

A DNA sequence (GBSx0781) was identified in *S.agalactiae* <SEQ ID 2253> which encodes the amino acid sequence <SEQ ID 2254>. Analysis of this protein sequence reveals the following:

35 Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 40 bacterial cytoplasm --- Certainty=0.2589 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10037> which encodes amino acid sequence <SEQ ID 10038> was also identified.

45 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2255> which encodes the amino acid sequence <SEQ ID 2256>. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

-832-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2767 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 80/86 (93%), Positives = 84/86 (97%)

Query: 6 LKTIKENMTFEELPLGLKAKKQVVRTGWGGAENYVQLFDITLEVNGKVLQATPYFLINVT 65  
 + IKENMTFEELPLGLKAKKQVVRTGWGGAENYVQLFDITLEV+GKVLQATPYFLI+VT  
 Sbjct: 3 ISSIKENMTFEELPLGLKAKKQVVRTGWGGAENYVQLFDITLEVNGKVLQATPYFLIHVT 62

Query: 66 GGGGFSMWAFTPCDVLAEDWIEVND 91  
 G GGGFSMWAFTPCDVLAEDWIEVND  
 Sbjct: 63 GAGGGFSMWAFTPCDVLAEDWIEVND 88

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 736

A DNA sequence (GBSx0782) was identified in *S. agalactiae* <SEQ ID 2257> which encodes the amino acid sequence <SEQ ID 2258>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA085256 GB:AB021978 3-oxoacyl-[acyl carrier protein] reductase  
 homolog [Moritella marina]

Identities = 82/239 (34%), Positives = 125/239 (51%), Gaps = 15/239 (6%)

Query: 2 TKVLVLTGCSAGIGYAQQYFLKQYQVYGVKSKDPENLN----GNFNP-IKLDSDDL 55  
 +K VLVTG + GIG A A++F K G V G S + G+ P ++L+++S  
 Sbjct: 5 SKTVLVTGASRGIGRAIAERFAKLGATVIGTATSQAERIGAYLGDAGAGLELVISQD 64

Query: 56 S-----PLFTMVPTVDILCNTAGILDAYKPLLEVSDELEHLFDINFFVTVRLTRHYLR 109  
 S + T V +DIL N AGI A L + ++E ++ D N RL + LR  
 Sbjct: 65 SVDALYARIKTQVGHIDILVNNAGIT-ADNIFLAKEDENCVIDTNLSLYRLCKPLR 123

Query: 110 RMVERKSGIILDMCSIASFTAGGGGAAYTSSKHALAGFTROLALDYAKDICIQIFGIAPGA 169  
 M+++ G IIN+ S+ GG A Y ++K L GT+ LA + A I + +APG  
 Sbjct: 124 GMMKQRHGRILINISGVVFTGNGGQANYAAKSGLLGFTKSLASEVASRGITVNAVAPGF 183

Query: 170 VQTAMTASFEPGGLAEWASSETPIGRWTKPSVARELTGFLASGKARSMDGELVKIDGG 228  
 ++T MTA E + + + P R +E+AE GFLAS A + GE + ++G  
 Sbjct: 184 ISTDWIARELTER--QKQTLLAQVPTSRIGSTTEIAETVGLASGASYYTGTHVNGS 240

There is also homology to SEQ IDs 2628 and 7170.

A related sequence was also identified in GAS <SEQ ID 9107> which encodes the amino acid sequence <SEQ ID 9108>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 206/232 (88%), Positives = 224/232 (95%)

Query: 1 MTKVVLVTGCASGIGYAQRYFLKQGYQVGVDSKPKMLNGNPNFIKLDLSSLSPLFT 60  
 10 Sbjct: 4 MTKVVLVTGCASGIGYAQRYFLKQGYQVGVDSKPKMLNGNPNFIKLDLSSLSPLFT 63  
 Query: 61 MVPTVDILCNTAGILDAYKELLVSDDEEHLDFINFFVTRITRHYLRMRVSKSGIII 120  
 +VP+VDILCNTAGILDAYKELLVSDDEEHLDFINFF TV+LTRHYLRMRVSK+SG+II  
 15 Sbjct: 64 VVPSVDILCNTAGILDAYKELLVSDDEEHLDFINFFATVKLTRHYLRMRVSKSGVII 123  
 Query: 121 NMCSIASFIAGGGGAAYTSSKHALAGFTQLALDYAKDQIIFGIAPGAVQTAMTANDFE 180  
 NMCSIASFIAGGGGAYTSSKHALAGFTQLALDYAKD I IFGIAPGAV+TAMTA+DFE  
 Sbjct: 124 NMCSIASFIAGGGGAYTSSKHALAGFTQLALDYAKDQIIFGIAPGAVKTAMTANDFE 183  
 20 Query: 181 PGLAEWVASETPIGRWTKPSEVSLTGTFLASKARSMQGEIVKIDGGWSLK 232  
 PGLA+WVA ETPIGRWTKP EVASLTGTFLASKARSMQGEIVKIDGGW+LK  
 Sbjct: 184 PGLADWVARETPIGRWTKPDEVASLTGTFLASKARSMQGEIVKIDGGWTLK 235

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9063> which encodes amino acid sequence

25 <SEQ ID 9064>. An alignment of the GAS and GBS sequences follows:

Score = 83.1 bits (202), Expect = 4e-18

Identities = 72/258 (27%), Positives = 106/258 (40%), Gaps = 36/258 (13%)

Query: 6 EVAFITGAASGIGKQIGETLKEKTFVVSFINQ-----KLDQVVADYTKGVDFAFVV 60  
 +V +TG ASGIG + LK+G V D++ ++ D++ P+V  
 30 Sbjct: 3 KVVLTGTCASGIGYAQRYFLKQGYQVGVDSKPKMLNGNPNFIKLDLSSLSPLFTNV 62  
 Query: 61 CDVTKKEADNAIDITVVEKIGRIDILVNAG-LQHVAIEDFTTKPFEMIKMLTAPFI 119  
 +DIL NAGL ++ E+ E+ I  
 35 Sbjct: 63 -----PTVDILCNTAGILDAYKELLVSDDEEHLDFINFFVTVTR 102  
 Query: 120 AIKRAFFTMAQKKGRIINNASINGVIGPAGKSAYNSAKHGLIGLTKVTALEAADSGITV 179  
 + M +K G IINN SI I G +N X+K L G T+ AL+ A I+  
 40 Sbjct: 103 LTRHYLRMRVSKSGIIINMCSIASFIAGGGGAAYTSSKHALAGFTQLALDYAKDQI 162  
 Query: 180 NALCPGYVDTPLVRGQGFEDLSKTRGIFLENVLEEVPLVPQKRLIDVQBIADYVSIVS 239  
 I PG V T + FE L E + P R E+A+ FLAS  
 Sbjct: 163 FGIAPGAVQTAMTANDFE-----PGLAEWVASETPIGRWTKPSEVSLTGTFLAS 212  
 45 Query: 240 DIKAGYTGQACILDDGYT 257  
 KA+ +G+ +DGG++  
 Sbjct: 213 GKARSMQGEIVKIDGGWS 230

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 2259> which encodes the amino

50 acid sequence <SEQ ID 2260>. An alignment of the GAS and GBS sequences follows:

Score = 427 bits (1086), Expect = e-122

Identities = 206/232 (88%), Positives = 224/232 (95%)

Query: 4 MTKVVLVTGCASGIGYAQRYFLKQGHVGVDSKPKMLNGNPNFIKLDLSSLSPLFT 63  
 55 MTKVVLVTGCASGIGYAQRYFLKQGHVGVDSKPKMLNGNPNFIKLDLSSLSPLFT 60  
 Sbjct: 1 MTKVVLVTGCASGIGYAQRYFLKQGHVGVDSKPKMLNGNPNFIKLDLSSLSPLFT 60  
 Query: 64 VVPSVDILCNTAGILDAYKELLVSDDEEHLDFINFFATVKLTRHYLRMRVSKSGVII 123  
 +VP+VDILCNTAGILDAYKELLVSDDEEHLDFINFF TV+LTRHYLRMRVSK+SG+II  
 60 Sbjct: 61 MVPTVDILCNTAGILDAYKELLVSDDEEHLDFINFFVTRITRHYLRMRVSKSGIII 120  
 Query: 124 NMCSIASFIAGGGGAYTSSKHALAGFTQLALDYAKDQIIFGIAPGAVKTAMTANDFE 183  
 NMCSIASFIAGGGGAYTSSKHALAGFTQLALDYAKD I IFGIAPGAV+TAMTA+DFE

-834-

5      Sbjct: 121 NMC8IASFVLAGGGGAATTSSKHALAGFTWQLAIDYAKDCIQIFGLIARGAVQTAMTASDFE 180

Query: 184 PGGLADWVARETFIGRWTKPDEVARLIGFLASGKARSMQGEIVKIDGGWTLK 235

PGGLA+WVA KTFIGRWTKP KVARLIGFLASGKARSMQGEIVKIDGGW+LK

5      Sbjct: 181 PGGLAEWVASETFIGRWTKPSEVARLIGFLASGKARSMQGEIVKIDGGWTLK 232

SEQ ID 2258 (GBS251) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 2; MW 21.7kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 6; MW 52kDa).

- 10      Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 737

A DNA sequence (GBSx0783) was identified in *S. agalactiae* <SEQ ID 2261> which encodes the amino acid sequence <SEQ ID 2262>. Analysis of this protein sequence reveals the following:

- 15      Possible site: 48  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL      Likelihood = -3.82      Transmembrane      62 - 78 ( 62 - 79)  
 20      ----- Final Results -----  
          bacterial membrane --- Certainty=0.2529(Affirmative) < succ>  
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 25      No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 738

A DNA sequence (GBSx0784) was identified in *S. agalactiae* <SEQ ID 2263> which encodes the amino acid sequence <SEQ ID 2264>. Analysis of this protein sequence reveals the following:

- 30      Possible site: 31  
 >>> Seems to have no N-terminal signal sequence  
 35      ----- Final Results -----  
          bacterial cytoplasm --- Certainty=0.1495(Affirmative) < succ>  
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 40      >GP:CAA20397 GB:AL031317 SC64.19c, unknown, len: 190 aa; contains  
          Pro-Ser- rich domain at N-terminus (Streptomyces  
          coelicolor A3(2))  
          Identities = 26/80 (32%), Positives = 44/80 (54%), Gaps = 5/80 (6%)  
 45      Query: 1      MDSDNEAICITETKVDIVPFKDVADHAFKEGEGDKLEWNRKAHIDFF----KPYFE 55  
          +DS + + +IE+T+V +VP +V HA EREGD + + WR H P + +  
          Sbjct: 103 VDSRERFVAVIEVTEVRVVLAEVDLAHAVDEGEGDTSVAGWRACHERFWHGAEMRAALG 162  
 50      Query: 56      EFGLMFSEDESRIVLEEFQVV 75  
          + G + + +VLE F+V  
          Sbjct: 163 DPGFTVDDATFVVLERPRIV 182



-835-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 739

- 5 A DNA sequence (GBSx0785) was identified in *S.agalactiae* <SEQ ID 2265> which encodes the amino acid sequence <SEQ ID 2266>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.49 Transmembrane 3 - 19 (3 - 19)
10
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
15

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BA06422 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 133/315 (42%), Positives = 191/315 (60%), Gaps = 4/315 (1%)
20
Query: 1 MKLAVLTGMIVKEVLPVLQKIGIDLVAILSTVRSLETAKDLAKENYMSLATSEYKAVL 60
 MK+A +GTG IV+ L L L I+G VA+ S R TAK LA +YN+ + + +L
Sbjct: 1 MKIATVGTGPIVEAFPSALDDIDGPMCMVMS--RKETIAKPLADQYNIPTIYTHFDHML 58

Query: 61 DNEEDITVYIGLPHILHFDYAKEALLAGKHVCEKPFPLEASQLELVSIANTROLILLE 120
 + ++ VI+ PN LH+ +A +AL KHVCEKPFPT A +LE L+S+A +L+L E
25
Sbjct: 59 ADPNVVEVYVASPNSLHYQHQAQLEHRKHVCEKPFPTSTARELHLISVARKNELALPE 118

Query: 121 AITNQTLPNFDLVKEHLSMLGDIKIVECNYSQYSSRYDAFKRGEIAPAFNPEMGQALRD 180
 AIT +LPN+ L+RE++ LG IK+++CHYSQYSSRYD F GE FNP GGAL D
30
Sbjct: 119 AITTHLPHYQLIKENIHKLGSIKMIQCNYSQYSSRYDRFLSGETPNVFNPAFSGQALMD 178

Query: 181 LNIYNLHLVIGLFGEPITAGYLPNIE-RGIDTSGVLVLDYGRFTVTCIGAKDCSAEVKST 239
 +N+YN+H V+ LFG P A Y N GIDTSGVLVL Y HF + C+G KD +
35
Sbjct: 179 INVYNIFHVNLPFGPPEAAHYIANQHANGIDTSGVLVLKYPHFISECVGCKDTQSHNFVL 238

Query: 240 IQGDKGSIALIGPTNTPMKISLTMQGESHVYQLNGDRHMRHDEIFVPEGIISNLDPKRA 299
 IQG+KG I + N + + ++ Q S + D ++ +E + +F++
Sbjct: 239 IQGNGYIHVENGAGCNCRNVKTYLDDQTSLSNAQTNMALLYYETRTFYE-MYQANFERC 297

40
Query: 300 AQALESHTTAKVLD 314
 + L +S +VM+V++
Sbjct: 298 YELLESYSHSVWRVME 312

```

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 719> which encodes the amino acid sequence <SEQ ID 720>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have a cleavable N-term signal seq.
50
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

55
Identities = 233/314 (74%), Positives = 269/314 (85%)

Query: 1 MKLAVLTGMIVKEVLPVLQKIGIDLVAILSTVRSLETAKDLAKENYMSLATSEYKAVL 60
 MKLAVLTGMIVKEVLPVLQKIGIDLVAILSTVRSLETAKDLAKENYMSLATSEYKAVL 60
 MKLAVLTGMIVKEVLPVLQKIGIDLVAILSTVRSLETAKDLAKENYMSLATSEYKAVL 60

```

Sbjct: 1 MKLAVLSTGMIVKEVLPVLQKIDGIDLVAITLSTVRSITAKDLAKAHMPLATSKYEAIL 60

Query: 61 DNEEDITVYIGLNNHLFPDYAKFALLAGKHVICKPPTLASQLEKLVSINANTROLITLR 120  
NEEDITVYIGLNNHLFP YAKFALLAGKHVICKPPT A +L+RLV IA R+LILLE

5 Sbjct: 61 GNSIEDTVYIGLNNHLFPAYAKFALLAGKHVICKPPTWTAGELDLRLVIAARRKRLILLE 120

Query: 121 AITNQVLNFWDLAKHLSMLGDIKIVRCNYSQYSSRYDAFKRGRIAPAFNPMQGGALRD 180  
AITNQVL N +KEHL LGDIKIVRCNYSQYSSRYDAFKRG+IAPAFNP+MQGGALRD

10 Sbjct: 121 AITNQVLSNFWPTIKRHLDLQGLGDIKIVRCNYSQYSSRYDAFKRGDIAPAFNPMQGGALRD 180

Query: 181 LNTYNHLVIGLFGPEPITAGVLNPIERGITDSGLVLDYGHFETVCIGAKDCSAEVKSTI 240  
LNTYN+H V+GLFG P T QYL N+B+GIDTSG+LV+DY FK VCIGAKDC+AB+KSTI

15 Sbjct: 181 LNTYNHFFVGLFGPEKPTVQYLANVKGIDTSGMLAWMDYBQFKVVCIGAKDCTARIKSTI 240

Query: 241 CGQKGSATILGPTNIMPKISLTMVQGSHTVQLANCRHRMIDEFVIFSGIISNLDPKRAA 300  
QG+KGS+A+LG TMT+P++ L+++G E V N HRN++EFV F +I DP++

20 Sbjct: 241 GQKGSALVLAGATNITLFGVQLSLKGRHPQVINHNKHHRMYEEFVAFROMIDQRDFEKN 300

Query: 301 CALEHSRTVMKVL 314  
CALEHSR VM VL+

30 Sbjct: 301 QALEHSRAVMVLE 314

SEQ ID 2266 (GBS342) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 10; MW 36.6kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 2; MW 61kDa).

GBS342-GST was purified as shown in Figure 226, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 740

30 A DNA sequence (GBSx0786) was identified in *S. agalactiae* <SEQ ID 2267> which encodes the amino acid sequence <SEQ ID 2268>. Analysis of this protein sequence reveals the following:

Possible site: 19  
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0499 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:CB12535 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 41/127 (32%), Positives = 63/127 (49%), Gaps = 11/127 (8%)

45 Query: 1 MISSIGQVMLYVSNVEASADFWKKNVGFERKEVKTQGDVVTYI-VAPKLDSEVSFVLHDK 59  
MI IG V +YV + + + FW KVG+ G + + + VAPK +E V++ K  
Sbjct: 1 MIKIQITVAVVYVEDQKAKQFWTEKVGFDLADHFGWPEASWLEVAPK-GASTRLVIYIPK 59

50 Query: 60 AIIAQMSPELDIATPLSILFRITDIDSTYQKLTAN--EVMNIP-IVDMGSMRVFFSDNEN 116  
A M + SI+FE DI TY+++ N E + P ++ G+ F DD  
Sbjct: 60 A----MMKSGECMKASIVFPCEDIPGTYRKMKINVEFLGEPNQMSWTFI--VQFDFEDG 113

Query: 117 NYFAIRE 123  
N P ++E

55 Sbjct: 114 NVFLLKE 120

No corresponding DNA sequence was identified in *S. pyogenes*.

-837-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 741

A DNA sequence (GBSx0787) was identified in *S. agalactiae* <SEQ ID 2269> which encodes the amino acid sequence <SEQ ID 2270>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3402 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04569 GB:AP001510 unknown conserved protein in others
[Bacillus halodurans]
Identities = 46/144 (31%), Positives = 83/144 (56%), Gaps = 10/144 (6%)

Query: 1 MVKALETYIVINGSGRAVDIFYKDVFOADLVNMMTWEEH--DENC--LEDKDLIINQL 56
M+ + Y++ +G+G+ A++FY+D A+++ + T+ ++ PN KDLI++A L
Sbjct: 1 MILTYNFMILMDGGQAAIEFYQDALNAEVITTCYTGDLPEQSPMASVNMKDLILHAHL 60

Query: 57 IFDGIIRIQISDENPD-----FVYQAGKIVTAIIVGSVEAREIYEKLLKKSACEVQLELQ 111
+ L ISD+ D F +G VT A+ +VE E+++KL +E+ L+
Sbjct: 61 KLGEVDLMLISDQCLOVDPERFFQHSQSPVTIALITNNVENTTEVFQRLASGGEEIA-PLB 119

Query: 112 ETPWSPAYANLVDQFVMMQISTE 135
+TF+SP Y + D+FG+ W +ST+
Sbjct: 120 KTFPSPLYGQVTDKFGITMHWSTQ 143
```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 742

A DNA sequence (GBSx0788) was identified in *S. agalactiae* <SEQ ID 2271> which encodes the amino acid sequence <SEQ ID 2272>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03784 GB:AP001507 UDP-N-acetylglucosamine pyrophosphorylase
[Bacillus halodurans]
Identities = 238/453 (52%), Positives = 322/453 (70%), Gaps = 1/453 (0%)

Query: 1 MSN-YAIIANGKGTSMKSDLKFMHVKVSGITMLRHVFRVSQAIEFSKIVTVIGHKAEIV 59
MSN ++AIIAAG+GTRMKS L KV+H V G M++HV V A+ +IVT+IGH A+ V
Sbjct: 1 MSNRFAVILAAAGQTRMKSLLKVLHVVQCKPMQHVVDVQVSLGDFEIVTIIIGHGADAV 60

Query: 60 RDVLGDKSEFVMQTEQIYTGHAIVMMARESLATSGHITLVIAGVPLITGRSLKNLIDPHV 119
+ LG++ + +Q EQLSTGHAV+ AE L +G T+Y+ GDPLT+T E++ +++ +H
```

Sbjct: 61 KSQLGERVSYALQEBQLTGHAVLAQESALGRRGVTVLQDPTLLTARTIDHVMVSYHE 120  
 Query: 120 NHGNVATILTADAANPFYGRIRIIRNSDDEVTKIVBQDANDPQQVKKEINTGTVFDMQS 179  
 + AT+LTA+ A+P GYGRIRI V +IVE KDA E+Q+ E+NTGTV VIM+ +  
 Sbjct: 121 EEQAKATVLTAEADPTCYGRIVRNDKGLVERIVBHKDASEEQTITEVNTGTVCFDNEA 180  
 Query: 180 LPEALKDINTNNAQGEYLLTDVIGFKPKAGKKGAYKLRDPSLGLVNDRVALATAEKVM 239  
 LF+ALK+ + NNAQGEYLL DVI I + G+V AYK +E+LGVNDRVALA IE+VM  
 Sbjct: 181 LFCALKEVGNNAQGEYLLPDVTQILQTKGKVAAYKTARVETLGVNDRVALAQEBQVM 240  
 Query: 240 RHRIARQHMVNGVTVVNPDSAYIDIDVEIGESVIEPNVLKQGTIKGKGTLLNGSYLVD 299  
 + RI M GVT ++P+ Y+ D IG++VI P + QGT IG+G +L +L  
 Sbjct: 241 KRRINEAMWRKGVTFIDPEQTVSPDATIGQDVTIYPTGMVLQGTITIGECVLGPHTELK 300  
 Query: 300 DAQVGNVTVITNSMVEESILSDGVTVGCVYAHIRPOTSLAKGVHIGNFVEVKSGIGENTK 359  
 D++GN + S+V S++ + V+GP+HIRP + + V IGNFVEVK S IG+ +K  
 Sbjct: 301 DSKIGNKTAVKQSVVHNSVEGSRVSGFPFHIRPASHIHDDVRIGNFVEVKSGTIGKESK 360  
 Query: 360 AGHLTYIGNAEVGCVDNFGAGTITVNYDQNKPKTEIGSNVFIGSNSTLIAPLEIGDNALT 419  
 A HL+YIG+AEVG VNF G+ITVNYDG+NEF T+I + FIG NS LIAP+ IG AL  
 Sbjct: 361 ASHLSYIGDAEVGSRVNFSCSSTITVNYDGNKFLTKIEDDAFIGCNSNLIAPVTIGKAL 420  
 Query: 420 TAAGSTITVNPIDISAIAGRGQVKNKEGYANKK 452  
 AAGSTIT+VP D+++I R RQ NKE Y KK  
 Sbjct: 421 IAAGSTITEDVPSDALSIARARQTNKEHYVTKK 453

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2273> which encodes the amino acid sequence <SEQ ID 2274>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0461 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 345/458 (75%), Positives = 398/458 (86%)  
 Query: 1 MSNYAIIAAGKGTIRMSDLPKVMHKVSGITMLEHVPRSVQALPESKIVTVIGHKAEIVR 60  
 M+NYAIIAAGKGTIRMSDLPKV+HKVSG+TLEHVPRSV+AI P K VTVIGH+ E+VR  
 Sbjct: 1 MTNYAIIAAGKGTIRMSDLPKVLHKVSGITMLEHVPRSVKALPESKIVTVIGHKSEIVR 60  
 Query: 61 DVLGDKSEFVMTBQLTGHAVMMAREELATSGKHTLVIAAGTPLITGESLKLIDPHVN 120  
 VL D+S FV QTBQLTGHAVMMAR +L +GHTLVIAAGTPLITGESLKLIDPHVN  
 Sbjct: 61 AVLAQDQAFVHQBQLTGHAVMMAREELATSGKHTLVIAAGTPLITGESLKLIDPHVN 120  
 Query: 121 HKNVATILTADAANPFYGRIRIIRNSDDEVTKIVBQDANDPQQVKKEINTGTVFDMQS 180  
 HKNVATILTA A +PFGYGRIRIIRNSDDEVTKIVBQDANDPQQVKKEINTGTVFDMQS  
 Sbjct: 121 HKNVATILTADAQPFYGRIRIIRNSDDEVTKIVBQDANDPQQVKKEINTGTVFDMQS 180  
 Query: 181 PEALKDINTNNAQGEYLLTDVIGFKPKAGKKGAYKLRDPSLGLVNDRVALATAEKVM 240  
 PEALK I TNNAQGEYLLTDV + IP+ +KVGAY LRDF+ESLGVNDRVALA AE VIM  
 Sbjct: 181 PEALKCITTNNAQGEYLLTDVVALFRANKKVGAYTLRDFNESLGVNDRVALAIAETVIM 240  
 Query: 241 RHRIARQHMVNGVTVVNPDSAYIDIDVEIGESVIEPNVLKQGTIKGKGTLLNGSYLVD 300  
 RI ++EMVNGVT NP++ YI+ DVEI + +IE NVTLG+T IG GT+LNG+Y+VD  
 Sbjct: 241 QRITQKHMVNGVTVFQNPETVYIESDVEIAPDVLBSNVTLKGETHIGSGTVLNGYTVVD 300  
 Query: 301 ACQVGNVTVITNSMVEESILSDGVTVGCVYAHIRPOTSLAKGVHIGNFVEVKSGIGENTK 360  
 +++G++ +TNSM+ E S+++ GVTGVPYAHIRPOT+L + VHIGNFVEVKSG IGE TKA  
 Sbjct: 301 SEIGNCVTVITNSMIESVLAAGVTVGVPYAHIRPOTTLIRKRVHIGNFVEVKSGHIGETKA 360  
 Query: 361 GHLTYIGNAEVGCVDNFGAGTITVNYDQNKPKTEIGSNVFIGSNSTLIAPLEIGDNALT 420  
 GHLTYIGNA+VG VN GAGTITVNYDQNK++T IG + FIGSNSTLIAPLE+GD+ALT

-839-

Sbjct: 361 GHLTYIGNAQVGSSVNVGAGTTIVNYDQNKYETVIGDHAFIGENSTLIAPLEVGDHAI/T 420

Query: 421 AAGSTITTNDVPIDIAIGRGQVNBKGYANKKPHPSQ 458

AAGSTI+ VPIDSTIAIGR RQV KBGYA + HHBP+

5 Sbjct: 421 AAGSTISKTVPIDIAIGRGRQVTKBQYAKRLAHHPR 458

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 743

- 10 A DNA sequence (GBSx0790) was identified in *S. agalactiae* <SEQ ID 2275> which encodes the amino acid sequence <SEQ ID 2276>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1366(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14293 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 92/177 (51%), Positives = 124/177 (69%), Gaps = 4/177 (2%)

Query: 4 EEKTIKQTVFDGQIIKVAVDVLENGLGQSKRELVPHGAVATLAVTPERKIVLVQY 63

25 EESTI ++ +F G++I + V+DVELENG SKRE+V H GAVA LAVT E KII+VRSQ+  
Sbjct: 5 EEKTIKQTVFDGQIIKVAVDVLENGLGQSKRELVPHGAVATLAVTPERKIVLVQY 63

Query: 64 RKAIIEGSIYEIPAGKLETSBGSKEEALRELEBETGYTG-NLEILYSFTTAIGFCNEKI 122

30 Sbjct: 64 RKPLERTIVEIPAGKLENG--EPETALRELEBETGYTAKLTKITAFYTSQGFADIV 121

Query: 123 VLYLATULQKVENPRQDDDEVLLELSYEDCKQWVEKGMIDAKTIILALQYGLK 179

35 Sbjct: 122 HVPLASELSVLSEKRELEDEFEVMEVTLDELAKLVESREVDARTAYAIQYQLK 178

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2277> which encodes the amino acid sequence <SEQ ID 2278>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

- 40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1120(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 45

An alignment of the GAS and GBS proteins is shown below:

Identities = 136/182 (74%), Positives = 153/182 (83%)

Query: 1 MDPEEKTINRQTVFDGQIIKVAVDVLENGLGQSKRELVPHGAVATLAVTPERKIVLV 60

50 Sbjct: 1 MKPEEKTINRQTVFDGQIIKVAVDVLENGLGQSKRELVPHGAVATLAVTPERKIVLV 60

Query: 61 KQYKRAIEGSIYEIPAGKLETSBGSKEEALRELEBETGYTG-NLEILYSFTTAIGFCNE 120

55 Sbjct: 61 KQYKRAIEGSIYEIPAGKLETSBGSKEEALRELEBETGYTG-NLEILYSFTTAIGFCNE 120

Query: 121 KIVLYLATULQKVENPRQDDDEVLLELSYEDCKQWVEKGMIDAKTIILALQYGLK 180

Sbjct: 121 KIVLYLATULQKVENPRQDDDEVLLELSYEDCKQWVEKGMIDAKTIILALQYGLK 180

Query: 181 GG 182  
 GG  
 Sbjct: 181 GG 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 744

A DNA sequence (GBSx0791) was identified in *S. agalactiae* <SEQ ID 2279> which encodes the amino acid sequence <SEQ ID 2280>. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -15.44 Transmembrane 70 - 86 ( 64 - 88 )  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.7177(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2281> which encodes the amino acid sequence <SEQ ID 2282>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -15.60 Transmembrane 65 - 81 ( 58 - 83 )  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.7241(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 39/89 (43%), Positives = 61/89 (67%), Gaps = 6/89 (6%)  
 Query: 1 MGKELLLTDMIERSNRGEKVGQTILDQETKIISTEDGMBQLITENGKHIYKSRRIENAK 60  
 MG+ELLTLD+IE++ R E ++ +TK+++ + ++ IYKSRRIENAK  
 Sbjct: 2 MGRPLLLTDITIEKRRMETFEPDDAVNSDTKVMTLPE-----KDDKARIYKSRRIENAK 55  
 Query: 61 RNEPQRKINLVLFILLILLALLFYAIFKL 89  
 R++ Q KLN++L ++L+A+L YAIF L  
 Sbjct: 56 RSQLQSKINVLITAVMLLIALLVYAIFFVL 84

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 745

A DNA sequence (GBSx0792) was identified in *S. agalactiae* <SEQ ID 2283> which encodes the amino acid sequence <SEQ ID 2284>. This protein is predicted to be pfs protein (pfs). Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.32 Transmembrane 56 - 72 ( 56 - 72 )  
 ----- Final Results -----

-841-

bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 The protein has homology with the following sequences in the GENPEPT database:

>GF:AAC22869 GB:U32801 pfs protein (pfs) [Haemophilus influenzae Rd]  
 Identities = 100/229 (43%), Positives = 144/229 (62%)

10 Query: 1 MKIGIIAAMEELKLLVENLEDSKQSTVLSNVYSGRYGHEHLVLVQSGVGKMSMSVA 60  
 MKIGI+ AM +E++L + D+++ V S V + G+ ++ L+QSG+GKV +A+  
 Sbjct: 1 MKIGIVGAMQEVILKSLMDRTETRVASAVIFEGKINGKDVALLQSGIGKVAALIGIT 60

Query: 61 ILVESFKVDALINTGSAGAVATGLAVGVVVDVLTLYHDVLTAFGYDYQGMSPQLYFH 120  
 L++ K D +INTGSAG VA GL VGD+V++D YHD D+TAFGY+ GQ+ P F  
 15 Sbjct: 61 ALIQLAKPDCVINTGSAGVAGKLGKVDIIVISDETRYHDADVTAFGTEKQLEPAPFL 120

Query: 121 SDKTFVSTFEAVLSKEEMISKVGLIATGDSFIAGQSKIDVIKGFHPQVLAVEMEGAAIAQ 180  
 SDR ++ K+ K GLI +GDSFI ++KI IK FP V VEME AIAQ  
 Sbjct: 121 SDKKLADLAQETAEKQGSVGRGLICSGDSFINSDEKIAQKADFPNVTGVEMERATAIAQ 180

20 Query: 181 AAQATGKPFVVVRAMSDTAANDANITFDPIIIRAGKRSQAQLMAFLKAL 229  
 A FVVVVRA+SD R++F+EF+ A K+S+ ++ + L  
 Sbjct: 181 VCYAFNVFPVVVRAISDGGDGRMSFEEFLPLAAKQSSALVLMIDRL 229

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2285> which encodes the amino acid sequence <SEQ ID 2286>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1245 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 An alignment of the GAS and GBS proteins is shown below:

Identities = 169/229 (73%), Positives = 189/229 (81%)

Query: 1 MKIGIIAAMEELKLLVENLEDSKQSTVLSNVYSGRYGHEHLVLVQSGVGKMSMSVA 60  
 40 MKIGIIAAMEEL LL+ NL D + VLS YY+GR+G+HEL+LVQSGVGKMSAM+VA 60  
 Sbjct: 1 MKIGIIAAMEELSLLANLLDAQEHQVLSKTYTGTGPGKHELILVQSGVGKMSAM+VA 60

Query: 61 ILVESFKVDALINTGSAGAVATGLAVGVVVDVLTLYHDVLTAFGYDYQGMSPQLYFH 120  
 ILVE FK AIIINTGSAGAVA+ L +GDUVVD LVYHDV TAFGY YGOM+ QPLY+  
 45 Sbjct: 61 ILVEFKVAIINTGSAGAVASHLAIGDVVVDRLVYHDVDTAFGYATQGMAGQLYTD 120

Query: 121 SDKTFVSTFEAVLSKEEMISKVGLIATGDSFIAGQSKIDVIKGFHPQVLAVEMEGAAIAQ 180  
 D FV+ F+ VL E+ +VGLIATGDSF+AGQ+KID IK F VLAVEMEGAAIAQ  
 Sbjct: 121 CDQPVAFIPKQVLKHEKINGQVGLIATGDSFVAGQKIDQIKTAFSDVLAVEMEGAAIAQ 180

50 Query: 181 AAQATGKPFVVVRAMSDTAANDANITFDPIIIRAGKRSQAQLMAFLKAL 229  
 AA GKPF+VVRAMSDTAANDANITFD+PIIRAGKRSQA IM FL+ L  
 Sbjct: 181 AAHTAGKPFVVVRAMSDTAANDANITFDPIIRAGKRSQAQLMAFLKAL 229

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 746

A DNA sequence (GBSx0793) was identified in *S.agalactiae* <SEQ ID 2287> which encodes the amino acid sequence <SEQ ID 2288>. This protein is predicted to be StOR. Analysis of this protein sequence reveals the following:

-842-

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

5       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3777 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

10   A related GBS nucleic acid sequence <SEQ ID 9405> which encodes amino acid sequence <SEQ ID 9406> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF81675 GB:AF232688 Slor [Streptococcus mutans]  
 Identities = 97/175 (55%), Positives = 134/175 (76%)

15   Query: 1   MSEMIKMSISQELIVKDKLGYLLTKQGLLVSDLYRKRLVSVFLVNHLYTTADDIHEE 60  
           +SEM+KK++ E L++KDK GY LTR+G ++ S LYRKRL+EVFL+NHLYTTAD+IHEE  
   Sbjct: 38   VSEMVKKLLLEDVLKDKQAGYLLTKKQGLASSLYRKRLIEVFLVNHLYTTADSIHEE 97

20   Query: 61   AEVLEHTVSTTFVDQLEKLLDFFQFCPHGGTIPKKGFLVEINQMILQISQLGTIVISR 120  
           AEVLEHTVS FV++L+X L++F+ CPHGGTIF+ G+ LVE + TL ++++G Y++ R  
   Sbjct: 98   AEVLEHTVSDVFERLDKFLYFKVCPHGGTIPQHGQQLVERIYRITLKGVTGAVYLLKR 157

25   Query: 121 VHDQFQLKYLEQHRLEHINDIETLTQIDPIAKTYHITNDENLIPERIAQIYV 175  
           V D+PQLKY+SQH L I D + L + D +A Y I + E L + +ASQIY+  
   Sbjct: 158 VQDAPQLKYNBQHHLKIGDELRLLLEYDAFAGYTIKKGQQLQTSVAVASQIYI 212

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2289> which encodes the amino acid sequence <SEQ ID 2290>. Analysis of this protein sequence reveals the following:

30       Possible site: 39  
       >>> Seems to have no N-terminal signal sequence

35       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2910 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 44/75 (58%), Positives = 59/75 (78%)

40   Query: 1   MSEMIKMSISQELIVKDKLGYLLTKQGLLVSDLYRKRLVSVFLVNHLYTTADDIHEE 60  
           +SEMIKMSIS+ IVKDK GY L +G ++V++LYRK RL+EVFL++ L Y ++H+E  
   Sbjct: 38   VSEMIKMSISQGVIVKDKAKGYLLKDKGYALVANLYKRLIEVFLIHQLGNTQEVHQE 97

45   Query: 61   AEVLEHTVSTTFVDQ 75  
           AEVLEHTVS +P+D+  
   Sbjct: 98   AEVLEHTVSDSFIDR 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 50   Example 747

A DNA sequence (GBSx0794) was identified in *S.agalactiae* <SEQ ID 2291> which encodes the amino acid sequence <SEQ ID 2292>. This protein is predicted to be undecaprenyl pyrophosphate synthetase (uppS). Analysis of this protein sequence reveals the following:

55       Possible site: 46  
       >>> Seems to have no N-terminal signal sequence



-843-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3569 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9435> which encodes amino acid sequence <SEQ ID 9436> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP: CAB13526 GB: Z99112 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 88/165 (53%), Positives = 118/165 (71%), Gaps = 4/165 (2%)

Query: 1 MNLFVKFFDKYVPELDKNNRVQVIGDTHKLPKATYDAMORACLRTHNSGLVNFALNY 60  
 M LP +F + Y+PEL + NV+V+IGD LP T A+++A T N G++LNFALNY  
 Sbjct: 100 MKLFEEFLNTLYPELVEENVQVRIIGDETALFAHTLRAIEKAVQDTAQNDGMILNFALNY 159

Query: 61 GGRSEITNAIKEIAQDVLEAKINPDDITEDLVANHLMTNSLPVLYRDPDLIRTSGEHLRL 120  
 GGR+EI +A K +A+ V E IN +DI E L + +LMT SL +DP+L+IRTSGE+RL  
 Sbjct: 160 GGRTEIVSAANKLAKEVKESSLNEDIDESLFTYLMATESL---QDFELLIRTSGEHLRL 215

Query: 121 SNFLPWQSAYSSEFYFTFVLNPFDPFKDELHKAIVDYNQRRHRRGCV 165  
 SNF+ WQ AYSEF FT VLNPFDK+D +A+ ++ QR RRFQ +  
 Sbjct: 216 SNFLMWQAYSEFYFTDVLNPFDPFKENHFLQALGEFQQGRRRFGGI 260

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2293> which encodes the amino acid sequence <SEQ ID 2294>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2073 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 125/165 (75%), Positives = 145/165 (87%)

Query: 1 MNLFVKFFDKYVPELDKNNRVQVIGDTHKLPKATYDAMORACLRTHNSGLVNFALNY 60  
 MNLFV FFDKYPV L +NNV++Q+IG+T +LP+ T A+ +TK N+G+LNFALNY  
 Sbjct: 85 MNLFVFTTFDKYVPLHNNVKKIQMIGETSRLPDTLAAIMAAIDTKTKRTGLTNFALNY 144

Query: 61 GGRSEITNAIKEIAQDVLEAKINPDDITEDLVANHLMTNSLPVLYRDPDLIRTSGEHLRL 120  
 GGR+EI+ A++ IAQDVL+AKLND DITEDL+AN+LMT+ LPVLYRDPDLIRTSGEHLRL  
 Sbjct: 145 GGRSEITSAVRFIQAQDVLDALNPDGIDITDLIANYLMDHLFVLYRDPDLIRTSGEHLRL 204

Query: 121 SNFLPWQSAYSSEFYFTFVLNPFDPFKDELHKAIVDYNQRRHRRGCV 165  
 SNFLPWQSAYSSEFYFTFVLNPFDPFK EL KAI DYN+R RRFQ V  
 Sbjct: 205 SNFLPWQSAYSSEFYFTFVLNPFDPFKALLKAIADYNRRRRRFQV 249

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 748**

A DNA sequence (GBSx0795) was identified in *S. agalactiae* <SEQ ID 2295> which encodes the amino acid sequence <SEQ ID 2296>. This protein is predicted to be phosphatidate cytidylyltransferase (cdsA). Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have a cleavable N-term signal seq.

-844-

```

INTEGRAL Likelihood = -8.65 Transmembrane 201 - 217 (194 - 222)
INTEGRAL Likelihood = -7.96 Transmembrane 175 - 191 (170 - 197)
INTEGRAL Likelihood = -5.89 Transmembrane 81 - 97 (74 - 99)
INTEGRAL Likelihood = -3.03 Transmembrane 26 - 42 (23 - 42)
5 INTEGRAL Likelihood = -2.92 Transmembrane 136 - 152 (135 - 153)
INTEGRAL Likelihood = -2.02 Transmembrane 49 - 65 (47 - 66)
INTEGRAL Likelihood = -0.64 Transmembrane 248 - 264 (248 - 264)

----- Final Results -----
10 bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:BAB06141 GB:AP001515 phosphatidate cytidyltransferase
 [Bacillus halodurans]
 Identities = 116/266 (43%), Positives = 172/266 (64%), Gaps = 6/266 (2%)

Query: 1 NKERVINGAVAIIFIPFIMGGLPPFQVGLLAMIQVSEILRMRIIFSEBGLAMIG 60
 NK+RV+ + +F+ F+V+GGLPF + ++A I +SELL+M+++ FS GA +++
20 Sbjct: 1 NMQRVVTAIIFGLVFLTFVVVGGGLPFMTFIIVVATIANSELLKMKIAPFSPMGAPSLP 60

Query: 61 AFVLTVPLDSYLSFLVDASLSAYGIVIFMILAGTVLNSNSYPEDAAPIASSPFVVGIG 120
 ++L +P D + +P + + I +L TVL N+++F+A F I 98 Y+G G
25 Sbjct: 61 MNMLLLPNDMPKVVIFDFTKVEIFITFIFLLLVLTNTFTTFDAGSVILSSAYIGY 120

Query: 121 FGNLVBARMA---GIDKVLALPFWATDIGNVMIGRQPGQRKLLPSVSNKTIETSSLOG 177
 F L+ +R G+ V LF++WATD GAY GR PG+ KL P +SNKTIETSS+GG
30 Sbjct: 121 FHFLLSREIPIEGLPLVFFLVFWATDGGAYFAGRAFGRKHLNPHISNPKTIETSSG 180

Query: 178 TASAVIAVFFNMLFDKVIYAPHSFLVMLVLVAIPISIFQGFGLDVESSIKRHFGVDSGKL 237
 I A+++ F S+ V L ++ + S+FGQ GLDVES++KRH+ WDSG +
35 Sbjct: 181 IILAVIIGSLFVWIMPLF---SSYGVALAVIVVASVFOGLDGLVESAKRHVAVKDSQTV 237

Query: 238 IPGHGGILDRFDSMIFVFVFMHFGGL 263
 +FGHGGILDRFDS+I+V PI+H L
40 Sbjct: 238 LPGHGGILDRFDSLIVYMPFILMLHL 263

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2297> which encodes the amino acid sequence <SEQ ID 2298>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -9.98 Transmembrane 175 - 191 (170 - 197)
INTEGRAL Likelihood = -8.97 Transmembrane 5 - 21 (4 - 42)
45 INTEGRAL Likelihood = -6.85 Transmembrane 201 - 217 (197 - 222)
INTEGRAL Likelihood = -6.53 Transmembrane 81 - 97 (79 - 99)
INTEGRAL Likelihood = -4.73 Transmembrane 49 - 65 (47 - 71)
INTEGRAL Likelihood = -3.40 Transmembrane 136 - 152 (135 - 153)
INTEGRAL Likelihood = -3.24 Transmembrane 26 - 42 (22 - 42)
50 INTEGRAL Likelihood = -1.17 Transmembrane 248 - 264 (248 - 264)

----- Final Results -----
bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:BAB06141 GB:AP001515 phosphatidate cytidyltransferase
[Bacillus halodurans]
Identities = 125/266 (46%), Positives = 177/266 (65%), Gaps = 6/266 (2%)

60 Query: 1 NKERVVWGQVAVAFVLPFLIIGNLPQLFVGVLIAMIQVSEILMKLEVPSPBQGVAMLA 60
 MK+RVV + +FL F++G LDF +F+ V+A I +SELLMK++ FS G F+L
Sbjct: 1 NMQRVVTAIIFGLVFLTFVVVGGGLPFMTFIIVVATIANSELLKMKIAPFSPMGAPSLP 60

```

-845-

Query: 61 AFVLAVPMHYLTFLPIDANVAFYSLMVFVFLAGTVLNSRAYSFDAAFPFIATSFYVIG 120  
 ++L +P D + +P V + + +F+L TVL ++ED+A F I +S +Y+G G  
 Sbjct: 61 MWMLLLENDMFKVVIPOPTKVEIPFIFLILLTLVLTNTFTYDRAFPVILSSAYIGV 120

5 Query: 121 FQHLINAR--LSGIDKVFALFVWATDIGAYLIGRQGRKLLPTVSPNKTIKESLGG 177  
 F L+ +R G+ VF LF+ +WATD GAY GR FG+ KL P +SPNKTIKESLGG  
 Sbjct: 121 FPHLLEREIFEIGLFLVFFVLFWIATDSGAYFAGRAPGKHLMWHISPNKTIKESLGG 180

10 Query: 178 IACAVLWSFIPMWIDRSVYAPHHLFVTLNLVLALESIFAQGDVLESALKRHFVGDGSKL 237  
 I AV++ +F I +++ + +++VA S+F Q GDVLESALKRHH+ VKDSG +  
 Sbjct: 181 IILAVIIGSLFVWI-MPLPSSYGVAVLVVA--SVFQQLGLDLESALKRHHVAVKDSIV 237

15 Query: 238 IPGHGILDRFDSMIFVFPIMHFLGL 263  
 +PGHGGILDRFDS+I+V PI+HL L  
 Sbjct: 238 LPGHGILDRFDSLIYVPIHLHLGL 263

An alignment of the GAS and GBS proteins is shown below:

Identities = 204/264 (77%), Positives = 243/264 (91%)

20 Query: 1 MKERVINGAVALAIFIPFLVMGSLPFOFLVGLLAMIQVSELLMRLEISFSEGALAMIG 60  
 MKERV+RG VA+ATP+PFL+G LFPQ VG+LAMIQVSELL+MRLE+SFSEG NM+  
 Sbjct: 1 MKERVINGGVAVALIFLFFLLIGNLFFQLFVGVLAMIQVSELLMKRLEVSFSEGVFMAIA 60

25 Query: 61 AFVLTFLDLSYSLFLEVDASLSAYGIVIFMILAGTVLNSNSYSFEDAAPFIASSFVVGIG 120  
 AFVL VF+D YL+FLP+DR+++ Y +++F ILAGTVLNS +YSF+DAAPFLA+SFYVIG  
 Sbjct: 61 AFVLAVPMHYLTFLPIDANVAFYSLMVFVFLAGTVLNSRAYSFDAAFPFIATSFYVIG 120

30 Query: 121 PQMLVSRMAGIDKVLALLFVWATDIGAYMIGRQGRKLLPVSFNKTIKESLGGIAS 180  
 FQ+L+ +R++GIDRV LALFVWATDIGAY+IGRQFG+RKLFP+VSPNKTIKESLGGIA  
 Sbjct: 121 FQHLINARLSGIDKVFALFVWATDIGAYLIGRQGRKLLPTVSPNKTIKESLGGIAC 180

35 Query: 181 AIVAFPFMLPDKTIYAPHSLFVLMVLVAIPSIQFGDLVESSTIKRHFVGDGSKLIPG 240  
 A++V+V FM+ D++VYAPH FL MLVLVA+PSIF QFGDLVES++KRFVGDGSKLIPG  
 Sbjct: 181 AVLVSFIPMWIDRSVYAPHHLFVTLNLVLALESIFAQGDVLESALKRHFVGDGSKLIPG 240

40 Query: 241 HGGILDRFDSMIFVFPIMHFLGL 264  
 HGGILDRFDSMIFVFPIMH FLGL  
 Sbjct: 241 HGGILDRFDSMIFVFPIMHFLGL 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 749

A DNA sequence (GBSx0796) was identified in *S. agalactiae* <SEQ ID 2299> which encodes the amino acid sequence <SEQ ID 2300>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have an uncleavable N-term signal seq

50 INTEGRAL Likelihood = -11.09 Transmembrane 2 - 18 ( 1 - 25)  
 INTEGRAL Likelihood = -9.39 Transmembrane 394 - 410 ( 390 - 415)  
 INTEGRAL Likelihood = -8.01 Transmembrane 181 - 197 ( 173 - 198)  
 INTEGRAL Likelihood = -2.97 Transmembrane 343 - 359 ( 342 - 360)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5437 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47948 GB:AF152237 Eep [Enterococcus faecalis]  
 60 Identities = 229/425 (53%), Positives = 298/425 (69%), Gaps = 9/425 (2%)

Query: 1 MLGILTFIILFGVIVVHEFGHFFYFAKKSGLVREFAIGMGPKIPSHIDKSGTITITRLL 60  
 M I+TFII+FG+V+VHEFGHFFYFAK+GILVREFAIGMGPKIP+H K-GTITITR+L  
 Sbjct: 1 MKTITITFIIIVFGILVHEFGHFFYFAKRAGILVREFAIGMGPKIPFAHKGKDTITITRLL 60

5 Query: 61 PLGGYVRMGAGWGDOKTEIKTGTPASITLAKBQIVTRINLSGKQLNLSLPINVTADIED 120  
 P+GGYVRMGAG G+D TEI G P S+ LN G V +IN S K S+P+ V +DE  
 Sbjct: 61 PIGGYVRMGAGWGDMEITITPGNPLSVELNAVGNVVKINTSKKVQLPHSIPMSVVDPLDK 120

10 Query: 121 KLITITGLV---LSSTKTYVHDHDTIIEEDGTETIRIAPLDVQYNASVWGRILINFAGM 177  
 +L I G V E Y VHDHDTIE DGTE+RIAPLD+Q+Q+A + R+TNFAGM  
 Sbjct: 121 ELFIKGYVNGNKEEETVYKVDHDTIIEEDGTETVRIAPLDVQPSAKLSQRLINFAGM 180

15 Query: 178 NNFILGLVVFILAFILQGGVDLSTMQV-RVSEKGPAAAGLKNDRILQIGSHKVSWE 236  
 NNFILG ++P F+QGV DL+TQ+ +V NCPAA AGLK ND+L I + K+ +E  
 Sbjct: 181 NNFILGPILFTLAVFLQGGVDTLNTQIGQVINGPAAAGLKNDRVLSINNKKIKYE 240

20 Query: 237 OLTPAAVEKSTRHLKKQKLAKIKSKEVVKINVKPKVDKSVI--IGIMPALKTSPKDK 294  
 T V+K+ EK ++ KE T+ + QV+K I +G+ P +KT K  
 Sbjct: 241 DFTTIVQKNP---EKPLTFVVERNGKEQLTVTEPKQVKEQTIGKGVVYPMKTLPSK 297

25 Query: 295 ILGGILKAWSEFPFRIINELKKIHAHSINKLGGFVALYQASQPAKNGFVTVLNLAKGIS 354  
 L+GG++ S +I L L L PS+NKGQPV ++ S +A+ G TV+ LM +S  
 Sbjct: 298 LMGGIQDTLNSSTQIFKALGSLTGPFLNLKGGFVNMFLSSEASNAGVSTVVLAWLS 357

30 Query: 355 INGLINSLIPALDGGKIVNMLIEARREPLKQETETITLGVNMLVLMIAVWDI 414  
 +NLGI+NL+P+PALDGGKIV+NI+E +R KP+ E E ITL G ++VLM+ VWDI  
 Sbjct: 358 NNLGIINLLIPALDGGKIVNMLIIBSVRGKIPISPEKEGITITLGGFVNMVLMVAVWDI 417

Query: 415 MRFF 419  
 R FF  
 Sbjct: 418 QRFF 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2301> which encodes the amino acid sequence <SEQ ID 2302>. Analysis of this protein sequence reveals the following:

35 Possible site: 26  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -11.41 Transmembrane 2 - 18 ( 1 - 25)  
 INTEGRAL Likelihood = -9.77 Transmembrane 394 - 410 ( 390 - 415)  
 40 INTEGRAL Likelihood = -9.61 Transmembrane 180 - 196 ( 173 - 201)  
 INTEGRAL Likelihood = -2.66 Transmembrane 347 - 363 ( 343 - 363)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5564 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AND47948 GB:AP152237 Bep [Enterococcus faecalis]  
 Identities = 230/427 (53%), Positives = 298/427 (68%), Gaps = 13/427 (3%)

50 Query: 1 MLGIITFIILFGVIVVHEFGHFFYFAKKSGLVREFAIGMGPKIPSHVDQGGTITITRLL 60  
 M IITFII+FGILV+VHEFGHFFYFAK+GILVREFAIGMGPKIP+H +GT YT+R+L  
 Sbjct: 1 MKTITITFIIIVFGILVHEFGHFFYFAKRAGILVREFAIGMGPKIPFAHKGKDTITITRLL 60

55 Query: 61 PLGGYVRMGAGWGDOKTEIKTGTPASITLNEQGVTRINLSGKQLNLSLPINVTADIED 120  
 P+GGYVRMGAG G+D TRI G P S+ LN G V +IN S+ P S+PM V +DE  
 Sbjct: 61 PIGGYVRMGAGWGDMEITITPGNPLSVELNAVGNVVKINTSKKVQLPHSIPMSVVDPLDK 120

60 Query: 121 QLSITGLV---LSSTKTYKVHDHDTIIEEDGTETIRIAPLDVQYNASVWGRILINFAGM 177  
 +L I G V EE YKV HDHDTIIE DGTE+RIAPLDVQ+Q+A + R+TNFAGM  
 Sbjct: 121 ELFIKGYVNGNKEEETVYKVDHDTIIEEDGTETVRIAPLDVQPSAKLSQRLINFAGM 180

Query: 178 NNFILGLVVFILAFILQGGVDPFSSNV-RVCRNGAAAGLKNDRQIVAINYKVTW 236  
 NNFILG ++F L VFLQGG+ D +N + +V NG AA+AGL+ND+++IN K+ +  
 65 Sbjct: 181 NNFILGPILFTLAVFLQGGVDTLNTQIGQVINGPAAAGLKNDRVLSINNKKIKYE 240

Query: 237 DLTEAVDLATRDIGPSQTIKVYKSHQRLKTVAVKQKH-AKTYTI---GVKASLKTGFK 292  
 D T V P + + + + + V P+K + TI GV +KT  
 Sbjet: 241 DFTTIV----QIMPEKPLTFVVERNGKEEQLTVPTEKQKVEKQTIGKGVVPMCTDLP 295

Query: 293 DKLLGGLRLAWSRAFTIINALKGLITGFSINKLGGPVAMVYMSNQANGLSLSLMM 352  
 KLG+G+ + I AL L TGFSSINKLGGPV M+ +S +A+ G+ +V+ LGAM  
 Sbjet: 296 SKLMGGTQDTLSTTQIFKALGSLFTGFSINKLGGPVMMFKLSSEASKAGVSTVPLVLM 355

Query: 353 LSNLGIINLPIPALDGGKILMIIEAIRRKPIKQSTEVYITLAGVAVVLMIAVTWN 412  
 LSNLGI NL+P+PALDGGKI++MIIE +R KPI E E ITL G ++VLM+ VTN  
 Sbjet: 356 LSNLGIINLPIPALDGGKIVLMIIEGVKGPISPEKGIITLIGFGVAVVLMIAVTWN 415

Query: 413 DIMRVFF 419  
 DI R FF  
 Sbjet: 416 DIQRPFF 422

An alignment of the GAS and GBS proteins is shown below:

Identities = 306/419 (73%), Positives = 359/419 (85%)

Query: 1 MLGIITFIIIFGIVVVHEFGHFYFAKSGILVREFAIGMGPKFISHIDEGTYYTIRIL 60  
 MLGI+TFIIIFG++V+VHEFGHFYFAKSGILVREFAIGMGPKFISH+D+ GT YT+R+L  
 Sbjet: 1 MLGIITFIIIFGLIVVHEFGHFYFAKSGILVREFAIGMGPKFISHVDQGGTLYTLNML 60

Query: 61 PLOGGVVMAGWGDDKTEIKTGTTPASLTINKEGIVTRINLGSQKLTSLP+INVAYDLED 120  
 PLOGGVVMAGWGDDKTEIKTGTTPASLTIN++G V RINLS +LD TSLP++VT IDLED  
 Sbjet: 61 PLOGGVVMAGWGDDKTEIKTGTTPASLTINLQGGFVIRINLSQSKLDP+SLPMHVITGDLED 120

Query: 121 KLITITGLVLETKTYVSDHDATIIREDGTEIRIAPLDMQVQNASVWGLRITNFGAPARNF 180  
 +L+ITGLVL ETKTY V HDATI+REDGTEIRIAPLD+QVQNAS+ GRLLITNFGAPARNF  
 Sbjet: 121 QLSITGLVLETKTYKVAHDATIVREDGTEIRIAPLDVQVQNASIGRLLITNFGAPARNF 180

Query: 181 ILGLVVFIALAFIQGGVQDLSTHWVRVSENGPAASAGLANNDRIQLGSHKYSWEGQLTA 240  
 ILG+VVFIL F+QGG+ D S+N VRV ENG AA AGL+ND+I+ I +KV++W IT  
 Sbjet: 181 ILGIVVFILLVFLQGGMPDFSSNHVRVQENGAAAGLNDRIQLGSHKYSWEGQLTE 240

Query: 241 AVEKSTHLEKKQKLKIKSEKVPKTNVKKPQVDSKSYIIGIMPAKTSFKDKLLGGLK 300  
 AV+ +TR L Q + + KS + +KT+ VKPQK K+Y IG+ +LKT FKDKLLGGL+  
 Sbjet: 241 AVDLATRDIGPSQTIKVYKSHQRLKTVAVKQKHAKTYITIGVKSALKTGFKDKLLGGL 300

Query: 301 LAWESFFRIINELKULIAHFSINKLGGPVALYQASSQAQKGFVTVLNLMLISINLGM 360  
 LAW F ILN LK LI FS+KLGGFVA+Y S+QA+NG +VL+LM ++SINLGI  
 Sbjet: 301 LAWRAFTIINALKGLITGFSINKLGGPVAMVYMSNQANGLSLSLMM+SLNLTG 360

Query: 361 NLIPIPALDGGKILNITIEAIRRKPKQETETTYITLAGVAVVLMIAVTWNIMRVFF 419  
 NLIPIPALDGGKI+NI+EAIRRKPKQETETTYITLAGVA+N+VLMIAVTWNIMRV FF  
 Sbjet: 361 NLIPIPALDGGKILNITIEAIRRKPKQETETTYITLAGVAVVLMIAVTWNIMRVFF 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 750

A DNA sequence (GBSx0797) was identified in *S. agalactiae* <SEQ ID 2303> which encodes the amino acid sequence <SEQ ID 2304>. This protein is predicted to be prolyl-tRNA synthetase (proS). Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.32 Transmembrane 473 - 489 ( 473 - 490)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-848-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 10181> which encodes amino acid sequence <SEQ ID 10182> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB13530 GB:Z99112 prolyl-tRNA synthetase [Bacillus subtilis]
Identities = 301/608 (49%), Positives = 410/608 (66%), Gaps = 52/608 (8%)

10 Query: 1 MKQSMILPTLEMPSDAQVISHALMVRAGYVRCVSAGIYAYLPLANKTIEKFKITMRQE 60
M+QS LIPTLER+P+DA+ SH L++RAG++RQ ++G+Y+I+PLA + I+ + I+R+E
Sbjct: 1 MRQSLTILPTLEVPADAEAKSHQQLLRAGFRIGNTSGVSYNPLAYKVIQNIQQIVREE 60

Query: 61 FEKIGAVEMLAPALLTADLWRESGRYETTYGDELTKLNQDSDFILGPTTHEETPTLAVRD 120
EKI AVEML PAL A+ W+ESGR+ TYG +L +LK+R +F LG THEE T+LVRD
15 Sbjct: 61 MEKIDAVEHMLPALQQAETWQESGRWYTYGPELMRLKDRHGREPALGATHEVITSLVRD 120

Query: 121 AVKSYKQLPILNLYQIQSKYRDEKFRPNGLLRTREFIMKQGYSHNDYEDLDVTEYDVKRA 180
VKSXK+LPL LYQIQSK+RDEKFRP GLLR REFIMKQ YSH E LD TY+ +A
Sbjct: 121 EVKSYKRLPILNLYQIQSKFRDKNRFPGLLRGREFIMKDATSFHASBSLDETYQWVEA 180

20 Query: 181 YEAFITRAGLDPKGIIGDGGAMGKDSQEFMAVTPNRTDLNRKLVLDKTIPISEDDIPEDV 240
Y IF R G++ + I D GANQGD+ EFMA++
Sbjct: 181 YSNI FARQGINVRVFIADSGANMGKDTHFMAIS----- 214

25 Query: 241 LEEIKVELSAWLVSQSDTIAYSTESSYAANLEHATNEYKPSKTAATFEBVTXVETPNCKS 300
GEDTIAYS ES YAAN+ENA ++ + + RV TEN KA+
Sbjct: 215 -----AIGKDTIAYSDESQYAANLEHATNEYKPSKTAATFEBVTXVETPNCKT 262

30 Query: 301 IDEVAGFLSIDENQTIKTLFLPADQEPVALLVQNDQVNDVKLYLAADFLEPASBQLQ 360
I+E+ FL + +LK+LF AD++ V+ L+ G+ +VND+K+VM L A+ +E A+ E+
Sbjct: 263 IEELTAPLQVSARACIKSVLPKADDRFVLVLVRGDHEVNDIKVQMLHAEVVELATHEEV 322

Query: 361 KEIFGAGQSLGPNVLPDSVKIADRKVKQDLANAVSGANQDGYHFTGVNPERDFTA-EYV 419
+ G G +GPV + V++ AD+ V+ + NAV+GAN+ +H+ VN RD E+
35 Sbjct: 323 IQLGTEPGFVPGVGIHQVEVYADQAVKAMVNAVAGNEGDHXYKVMVNRDAQIKEIFA 382

Query: 420 DIREVKEGETSPDGKGTILKFGARGIEIGHIFKLGTRYSDMGANILDEKGRSNPIVMGCGY 479
D+R +KEG+ SPDGKGT++FA GIE+G +FKLGTRY++M A LDENGR+ P++MGCGY
40 Sbjct: 383 DLRFKEGDSPDGKGTIRFAGIEVQGVFKLGTRYSEMANATYLDENGRAPMLMGCGY 442

Query: 480 IGVSRLGAVTEQHARLFPVNKTPKGAIRFAMGNIPEELAPFVHLITLVNVDQESQDLT 539
IGVSR LSA+ EQH G+ +P+ +AP+D+H++ +N+K+ ++L
Sbjct: 443 IGVSRTLSAIAEQH-----HDEKGLWPKSVAPYDLHILALN+NDQGRILA 489

45 Query: 540 EKIEADLMLKGYEVLTDORNERVSGSKSDSDGLGPIRVITVGKKASGIVGVKIKASGDT 599
EK+ ADL +GYEVL DOR ER G KP+SDSDGLGPIR+TVGK+A RGIVE/KI+ +G++
Sbjct: 490 EKLAYDLKABGYEVLDDRAERAGVKPDAEDLGLPIRITVGKRADEGIVEKICRTQBS 549

50 Query: 600 IEVHADNL 607
E+ D L
Sbjct: 550 TEISVDL 557

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2305> which encodes the amino acid sequence <SEQ ID 2306>. Analysis of this protein sequence reveals the following:

```

55 Possible site: 16
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.32 Transmembrane 473 - 489 (473 - 490)

----- Final Results -----
60 bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 535/617 (86%), Positives = 584/617 (93%)

```

5 Query: 1 MQQSKMLIPITLREMPDQAQVISHALMVRAGTVRQVSGIYAYLPLANKRTIEKFTIMRQE 60
 Sbjct: 1 MQQSKLLIPITLREMPDQAQVISHALMVRAGTVRQVSGIYAYLPLANKRTIEKFTIMREE 60

 Query: 61 FEKIGAVEMLAPELLADLWRESGRYETYGEDLYKLKNDQSDPILGPHTEHTPTTLVRD 120
 Sbjct: 61 FEKIGAVEMLAPELLADLWRESGRYETYGEDLYKLKNDQSDPILGPHTEHTPTTLVRD 120

 Query: 121 AVKSYKQLPINLYQIQSKYRDEKRPNGILLRTREFIMKDGYSFHKQYEDLDVTEYDYKA 180
 Sbjct: 121 AVKSYKQLPINLYQIQSKYRDEKRPNGILLRTREFIMKDGYSFHHNVEDLDVTEYDYRQA 180

15 Query: 181 YEAIFTRAGLDFKGIIGDQGWGKDSQEFMAVTPNKTDLNRMLVLDKTIPISDIDIPDV 240
 Sbjct: 181 YEAIFTRAGLDFKGIIGDQGWGKDSQEFMAVTPRTDLRW+VLDK+I S+DIP+V

 Query: 241 LEEIKVELSAMLVSGEDTIAYSTESSYAANLEMATNEYKPSKKAATFEVTKVETPNCKS 300
 Sbjct: 241 LEDIKAEAAWMSGEDTIAYSTESSYAANLEMATNEYKPSKVAEADALAEVTEPHCKT 300

 Query: 301 IDEVAGFLSIDENQTIKTLFLADBPVALLVGNQVNDVCLKNYLAADFLPASEEQA 360
 Sbjct: 301 IDEVAFLSVDEVTQIKTLFLVADNEPVALLVGNCHINTVCLKNYLAADFLPASEEZA 360

25 Query: 361 KEIFGAGPGLSPVNLPSVKIADRKVQDLANAVSGANQDGYHFTGVNPERDFTAAYVD 420
 Sbjct: 361 RAFTGAGPGLSPVNLPSQSRIVADRKVQDLNVAAGANKDGFHMTGVNPGRDFOAYVD 420

 Query: 421 IREVKEGEISPDGKGLTFKARGIEIGHIFKLQTRYSDSMGANILDENGRNPVIMCKGI 480
 Sbjct: 421 IREVKEGEISPDGKGLTFKARGIEIGHIFKLQTRYSDSMGANILDENGRNPVIMCKGI 480

35 Query: 481 GVSRIISAVIEQHARLFVNKTPKGYRFAMGINFPELAPFDVHLITVNVKQSDQDTE 540
 Sbjct: 481 GVSRIISAVIEQHARLFVNKTPKGYRFAMGINFPELAPFDVHLITVNVKQDQVQDTE 540

 Query: 541 KIEADLMLKGYEVLTDNRNERSGKFSDSDLIGLPIRVTVGKKAAGEVIEIKIAGDIT 600
 Sbjct: 541 KIEADLMLKGYEVLTDNRNERSGKFSDSDLIGLPIRVTVGKKAAGEVIEIKIAGDIT 600

45 Query: 601 EVHADNLIETLEILTKE 617
 Sbjct: 601 EVHADNLIETLEILTKE 617

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 50 Example 751

A DNA sequence (GBSx0798) was identified in *S. agalactiae* <SEQ ID 2307> which encodes the amino acid sequence <SEQ ID 2308>. This protein is predicted to be peptidoglycan hydrolase (flgJ). Analysis of this protein sequence reveals the following:

```

55 Possible site: 21
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.86 Transmembrane 9 - 25 (9 - 25)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.1744 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
60 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

-850-

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]  
 Identities = 101/201 (50%), Positives = 122/201 (60%), Gaps = 9/201 (4%)

5 Query: 2 KSRKKDLVLRLLT-----TLVPLG---GGVWFYNYKNDVETVTSASDQTTTFIQ 52  
 KS+KK K VL +L+ GL G + N+ +E +T +T FI  
 Sbjct: 16 KSKKKKSVLLFPKFPQKWSLIFGLFSLGLLASLNFPRIMEKNMTPTDSTVAFIAR 75

10 Query: 53 ISPTAIEISRTYDLVASVLLAQAILLESSSQGSDLSKAPNYNLPGLIKGEYKGSVQMPFLR 112  
 I T+ ++ DLVASV++AQAILLES SQGS LS+ P YN FGIKGEY G+S+ +T E  
 Sbjct: 76 IGTSRYLAERNDLVASVMIQAAILLES SQGSQLQKPLYNFPGIKGEYMGQSVLPFW 135

Query: 113 DDGKGNMTIQAPFRAYPNYSASLYDYAELVSSQKYASVWKSNTSSYKDATALGLVAT 172  
 DDGKGN I A FR+Y + SL DY E + Y V +S T SYKDATALGL+YAT  
 15 Sbjct: 136 DDGKGNFYHDAAPRSYGSVENSLOQVFELEGSYTVGVHRSKTRSYKDATALGLVAT 195

Query: 173 DTAYASKLNQIETYSLOYD 193  
 DT Y KLN IIR Y L YD  
 Sbjct: 196 DTTYGKLNIIIRYQLTIYD 216

20

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2309> which encodes the amino acid sequence <SEQ ID 2310>. Analysis of this protein sequence reveals the following:

Possible site: 24

25 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CA94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]  
 Identities = 103/189 (54%), Positives = 126/189 (66%), Gaps = 4/189 (2%)

35 Query: 4 KKKKLVLSLFLVLAACLGAYSAMRQSHKTSNVSARTLASSSTRHFIDEIGTASTIGQER 63  
 +K L+ I LF L L + + R + + + T +T FI BIG T+ +  
 Sbjct: 32 QKWSLIFGLFSLGLLASLNFPRIMEKNM-----TPIDETTVAFIARIGTSRYLAARN 87

40 Query: 64 DLVASVMIQAAILLESSNGKSSLSQAPYNNFPGIKGAYNGSSVTMTWEDGKNNTYTDQ 123  
 DLVASVMIQAAILLES +G+S LSQ P YNFFGIK YNG SVT+ TWEDG GN Y ID  
 Sbjct: 88 DLVASVMIQAAILLESSSQGSQLQKPLYNFPGIKGEYMGQSVLPFWTWEDGKNFYHDA 147

Query: 124 AFRAYPSIADSLNDYADLLSSSTYIGARKNTLSVQDATAALGLVATDTSYNLKNHII 183  
 AFR+Y S+ +SL DY + L S Y+G +S T SY+DATAALGL+YATDT+Y KLN+II  
 45 Sbjct: 148 AFRSYGSVENSLOQVFELEGSYTVGVHRSKTRSYKDATALGLVATDTSYNGKLNHII 207

Query: 184 ATYGLTAYD 192  
 Y LT YD  
 50 Sbjct: 208 EQYQLTIYD 216

An alignment of the GAS and GBS proteins is shown below:

Identities = 108/192 (56%), Positives = 124/192 (64%), Gaps = 2/192 (1%)

55 Query: 3 SRKKDLVLRLLTITLLVFGIGGVWFYNYKNDVETVTSASDQTTTFIQISPTAIEIS 61  
 ++KK KLVLR L G ++K NV T AS T FI I PTA I  
 Sbjct: 2 TKKKKLVLSLFLVLAACLGAYSAMRQSHKTSNVSAR-TIASSSTRHFIDEIGTASTIG 60

60 Query: 62 RTYDLVASVLLAQAILLESSSQGSDLSKAPNYNLPGLIKGEYKGSVQMPFLRQKGNWTO 121  
 + DLVASV++AQAILLES+G+S LS+AP YN FGIKG Y G SV M T EDG GN  
 Sbjct: 61 QERDLVASVMIQAAILLESSNGKSSLSQAPYNNFPGIKGAYNGSSVTMTWEDGKNNTYTT 120

Query: 122 IQAPFRAYPNYSASLYDYAELVSSQKYASVWKSNTSSYKDATALGLVATDTAYASKLN 181



-851-

I FRAYP+ + SL DYA+L+SS Y KSNT SY+DATAAL/GLYAITT-Y KLN  
 Sbjct: 121 IDQAFRAYPSIADSLNDYADLLSSSTYIGARKENTLSYQDATAAL/GLYAITTYSYNKLN 180

Query: 182 QIIETYSLDAID 193  
 II TY L AYD  
 Sbjct: 181 NIATYGLTAYD 192

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9073> which encodes the amino acid sequence <SEQ ID 9074>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 130 bits (323), Expect = 2e-32  
 Identities = 68/169 (40%), Positives = 96/169 (56%), Gaps = 3/169 (1%)

Query: 30 MMTLKLGNORLAFY---ADHETLTFRKI SHRAQSVQKKOLYSSVMMQAQILSHNNGS 86  
 +W N + P A +T TF++ IS A +++ LY+SV++AQAILSS+G+S

Sbjct: 25 VMFYNTKNDXNVEPTVSASDOTTTFTIQTISPTAIEISKTYDLASVLLAQAILSSSGQS 84

Query: 87 QLSQKPYTNFFGIKGSYKERSVIFPTLEDDGQNLVQIDAAFSGSLTACFLDYARVLN 146  
 Ls+ P YN FGIGK YK +SV PILEDG+GN+ QI A FR+Y +A DYA +++

Sbjct: 85 DLSKAPYNFLGIGKEYGKGSVQMPITLEDGKGNMTQIQAPFRAYPNYSASLYDAELVS 144

Query: 147 DFLYDKTHKFWSHYQXXXXXXXXXXXXXXXXXXXXKLNELIEWYQLTNFD 195  
 Y K S Y+ KLN++IE Y L +D

Sbjct: 145 SQKASVVKENTSSYDATALGLYATDTAYASKLNQIIETYSLDAID 193

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9075> which encodes the amino acid sequence <SEQ ID 9076>. An alignment of the GAS and GBS sequences follows:

Score = 69.1 bits (166), Expect = 1e-13  
 Identities = 52/151 (34%), Positives = 79/151 (51%), Gaps = 10/151 (6%)

Query: 2 TFLDKIKQQLDGNWAKYKILPSLTAQAILESQWKEH---APINLFGIKADSSWTGKS 57  
 TF+ I ++ Y + S+ AQRILES G+ AP+ LFGIK + + GKS

Sbjct: 48 TFIQTISPTAIEISKTYDLASVLLAQAILSSSGQSLSKAPYNFLPQIKGE--YKGS 105

Query: 58 FDIKTQEEYQAGVVDIVDRFRAYDSWDESIAHQSGFLVNDPRYEA--IGETDYKACY 115  
 T E+ G +T I PRAY ++ S+ D+ + LV +Y +V + YK A

Sbjct: 106 VQMPITLEDGKGNMTQIQAPFRAYPNYSASLYDYAR-LVSSQKYASVVKENTSSYDATAL 164

Query: 116 AIXAGYATASSYVELLIQIIRKNDLQSWDR 146  
 A+ YAT ++Y L Q+IE L ++D+

Sbjct: 165 ALTGL-YATDTAYASKLNQIIETYSLDAID 194

SEQ ID 2308 (GBS275) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 4; MW 22.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 4; MW 47.5kDa).

The GBS275-GST fusion product was purified (Figure 208, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 276), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 752

A DNA sequence (GBSx0799) was identified in *S. agalactiae* <SEQ ID 2311> which encodes the amino acid sequence <SEQ ID 2312>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.16 Transmembrane 876 - 892 (876 - 892)

----- Final Results -----
bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2313> which encodes the amino acid sequence <SEQ ID 2314>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.16 Transmembrane 873 - 889 (873 - 889)

----- Final Results -----
bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP: CAB94815 GB: A7245582 peptidoglycan hydrolase (Streptococcus thermophilus)
Identities = 96/202 (47%), Positives = 127/202 (62%), Gaps = 10/202 (4%)
```

```
Query: 4 KKRRRAKSSV-----NELVLGLN-LNLIVSMWTLGLNQRLAPYADHSTLTFRV 53
KK +++ KS + + + +GL LL L+ S+ +L ++ D T+ F+
Sbjct: 15 KSKKKKSVLLFPKFFQKWSLIFGLFSLGLGLLASLNFRLTMEKNMTPTDEITVAFIA 74

Query: 54 KISHAAQVAQKQKQLYSSVMMAQAILESNNKGSQLSQKPPYNNFPGIKGSYKERSVIFPTL 113
+I ++ +A + LY+SVM+AQAILES++G+SQLSQKP YNFFGIKG Y +SV PT
Sbjct: 75 RIGETSKYLAARNDLVAFVMAQAILESIDSQSQSQKPLNFPNGIKGYNGQSVTLPTW 134

Query: 114 EDDGQGNLYQIDAAPRSYGSILTACFLDYARVLNDPLYDKTHKFWSHYQDATAITLGTVA 173
EDDG+GN Y IDAAPRSYGS+ DY L Y H+ Y+DATA LIG YA
Sbjct: 135 EDDGKGNPYHIDAAPRSYGSVENSILQYVEFLEGSYYGVHRSKTRSYKDATAALGTVA 194

Query: 174 TDTTYHTKLNELIEWYQLTNFD 195
TDTTY KLN +IE YQLT +D
Sbjct: 195 TDTTYGDKLNSIIEYQLTIYD 216
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 1244/1468 (84%), Positives = 1351/1468 (91%), Gaps = 3/1468 (0%)

Query: 1 MSLEFKKIMDQIEMPLEIKNSSVFSSADIEVKVHSLSRIMEFHFSFPELLEFVYREIQ 60
MS+LF KIMDQIEMPLE+++ SS FSSADIEVKVHSLSRIMEFH+ F +LEI YREL
Sbjct: 1 MSDFAKIMDQIEMPLEMDRRSSAFSSADIEVKVHVSRLMEFHFAFAVLEIATYRELH 60

Query: 61 TRLVNSFEKADIKATFDIRARTIDFSDLLQYVYQQAFCPLCNSASGFSQSFLKWHYN 120
RL+ +FE ADIK TFDI+A +D+SDLLQ YYQ+AF CNSASGFSQSFLKWHYN 120
Sbjct: 61 DRLIRTFEADIKVTFDIQAQVQVYSDLLQAYYQAFHFAFCNSASGFSQSFLKWHYN 120

Query: 121 GSCMIISAPOFVNNHFRQNLPLEQCFSLGFGKLAIDMVSDEQMTQDLKSSFEFIRE 180
++II+AP FVNN+HFR NHLP L +Q FGFG L IDMVSD++MT+ L +F ++R+
Sbjct: 121 DDKGIIAARGFVNNHFRQNLPLEQCFSLGFGKLAIDMVSDEQMTQDLKSSFEFIRE 180
```

|    |             |                                                               |      |
|----|-------------|---------------------------------------------------------------|------|
| 5  | Query: 181  | QLLEKANQRAMQALSAQKSLSDGAPPSBVSPTQNYDFKERIKQRQAGFEKABITPMIEV   | 240  |
|    | 181         | L++KA Q+ LEAQKSLD PP ER TP ++D+KER ++RQAGFEKA ITPMIEH         |      |
| 10 | Query: 241  | TEENRIVPBGWVFSVERKTTTRTGRHIIINFKMTDYSSTFAMQKWAQKDEELTKYIMIKG  | 300  |
|    | 238         | TEENRIVPBGWVFSVERKTTTRTGRHIIINFKMTDYSSTFALQKWAQKDEELRQPMIAG   | 297  |
| 15 | Query: 301  | SWLPVKGNIENRINFTKSLTMMVQDIKEIVHERRKDLMPADQKRVFHTNMSTMDALPT    | 360  |
|    | 298         | SWLPVKGNIENRINFTKSLTMMVQDIKEIVHERRKDLMPBQKRVFHTNMSTMDALPT     | 357  |
| 20 | Query: 361  | VESLIDTAARWGHPALAITDHANVQSPFHGYHRACKAGIKATPLGANIVEDKVPISYNE   | 420  |
|    | 358         | VESLIDTAARWGHPALAITDHANVQSPFHGYHRACKAGIKATPLGANIVEDKVPISYNE   | 417  |
| 25 | Query: 421  | VDMNLHEATYVVDVETTGLSAANNLDLQIAASIKMFKGNIHQDFEIDPGHPLSAFTTE    | 480  |
|    | 418         | VDMNLHEATYVVDVETTGLSAANNLDLQIAASIKMFKGNIHQDFEIDPGHPLSAFTTE    | 477  |
| 30 | Query: 481  | LIGITDNRVSGSKPLQVLQBFQNFQCGTIVLVAHNIATPDVGFNMANYERHNLPLTQPIVI | 540  |
|    | 478         | LIGITDNRVSGSKPLQVLQBFQNFQCGTIVLVAHNIATPDVGFNMANYERHNLPLTQPIVI | 537  |
| 35 | Query: 541  | DTLEFARNLYPEYKRGHGLGPIKTRFQVLAHEHHHMANYPDARATGRLLPFLKARENRDVT | 600  |
|    | 538         | DTLEFARNLYPEYKRGHGLGPIKTRFQVLAHEHHHMANYPDARATGRLLPFLKARENRDVT | 597  |
| 40 | Query: 601  | NLMEINTKLVASDSYKARIKHATIVYQNGVGLQNFKLVSLSNLYKFGVARIPEVLD      | 660  |
|    | 598         | NLMEINTKLVASDSYKARIKHATIVYQNGVGLQNFKLVSLSNLYKFGVARIPEVLD      | 657  |
| 45 | Query: 661  | AHREGLLGTACSDGEVFDALLSNGIDPAVTLAKYDFIEVMPAIVPLVRLDLKBEV       | 720  |
|    | 658         | AHREGLLGTACSDGEVFDALLSNGIDPAVTLAKYDFIEVMPAIVPLVRLDLKBEV       | 717  |
| 50 | Query: 721  | GICQIIRDLIEVGRRLDKPVLATQNVHYLEPEDEIYRSIIIVRSLGQGMINRTIGRGA    | 780  |
|    | 718         | GICQIIRDLIEVGRRLDKPVLATQNVHYLEPEDEIYRSIIIVRSLGQGMINRTIGRGA    | 777  |
| 55 | Query: 781  | QPAFLPKAHFRTNEMLDEFALGKDLAYRIVVNTNINFADRFEDVEVKGDLTPFDVR      | 840  |
|    | 778         | QPAFLPKAHFRTNEMLDEFALGKDLAYRIVVNTNINFADRFEDVEVKGDLTPFDVR      | 837  |
| 60 | Query: 841  | AEEVVAELTYAKAFBIYGNPLPDIDIRIEKELSLILNGFAVITYLASQMLVQRNBERGY   | 900  |
|    | 838         | AEEVVAELTYAKAFBIYGNPLPDIDIRIEKELSLILNGFAVITYLASQMLVQRNBERGY   | 897  |
| 65 | Query: 901  | LVGSRGSGVSSFVAIMIGITEVNMPPHYVCPNQHSFPTDGSQSGSDLPNKNCPKQG      | 960  |
|    | 898         | LVGSRGSGVSSFVAIMIGITEVNMPPHYVCPNQHSFPTDGSQSGSDLPNKNCPKQG      | 957  |
| 70 | Query: 961  | TLYKKDQGDIPFETPLGFDGKVDPIIDLNPSGDDQPSMHLVDRIIPGEBYAFRAGTIVGT  | 1020 |
|    | 958         | TLYKKDQGDIPFETPLGFDGKVDPIIDLNPSGDDQPSMHLVDRIIPGEBYAFRAGTIVGT  | 1017 |
| 75 | Query: 1021 | AEKTAFGPVKGYERDYNKFYNDASVERKLATGAAGVKKRTQGFQGGIVIPINMYMDVIDFT | 1080 |
|    | 1018        | AEKTAFGPVKGYERDYNKFYNDASVERKLATGAAGVKKRTQGFQGGIVIPINMYMDVIDFT | 1077 |
| 80 | Query: 1081 | VQYPADDMTAANQTHFNPHDIDENVLKLDILGHDDPTMIRKLQDLGSDIPSNILPDDPD   | 1140 |
|    | 1078        | VQYPADDMTAANQTHFNPHDIDENVLKLDILGHDDPTMIRKLQDLGSDIPSNILPDDPD   | 1137 |
| 85 | Query: 1141 | VMKLFSGTVEVLGVIREQIGTPTGMLGIPRGTFNIVRGVMNETHPTTFAELQLSGLSHGT  | 1200 |
|    | 1138        | VMKLFSGTVEVLGVIREQIGTPTGMLGIPRGTFNIVRGVMNETHPTTFAELQLSGLSHGT  | 1197 |

Query: 1201 DVMILGNAQDLIKKGIATLSTVIGCRDDIMVYLMHAGLQPKMAFTIMERVRKGLNLIKISED 1260  
 DVMILGNAQDLIKKGIATL TVIGCRDDIMVYLMHAGL+PKMAFTIMERVRKGLNLIKISE+  
 Sbjct: 1198 DVMILGNAQDLIKKGIATLSTVIGCRDDIMVYLMHAGLQPKMAFTIMERVRKGLNLIKISEE 1257

Query: 1261 ERNGYIQAMRDNNVDPWYIISCGKIKYMFPPKAHAAAYVLMALRVAYFKVHPYFYCYAF 1320  
 ERNGYI AMR+NNVDPWYIISCGKIKYMFPPKAHAAAYVLMALRVAYFKVH+PI YYCYAF  
 Sbjct: 1258 ERNGYIDAMRENNDVDPWYIISCGKIKYMFPPKAHAAAYVLMALRVAYFKVHPYFYCYAF 1317

Query: 1321 STRAKAFELRTMSAGLDVAKARMKDITEKQRNEATNVENDLFTTLELVNEMLERGPKFG 1380  
 STRAKAFEL+TMS GLDVAKARM+DIT KR+ NEATNVENDLFTTLE+VNEMLERGPKFG  
 Sbjct: 1318 STRAKAFELKTMGGGLDVAKARMEDITIKRKNNEATNVENDLFTTLELVNEMLERGPKFG 1377

Query: 1381 KLDLYRSHATDFTIEEDTLIPPFVAMEGLGENVAQIVRAREDEGEFLSKTELKRGVSS 1440  
 KLDLY+S A +F I+ DTLIPPF+A+EGLGENVAQIV+AR++GEFLSK ELKRGV SS  
 Sbjct: 1378 KLDLYKSDAIRFQIKGDTLIPPFIALEGLGENVAQIVKARQGEFLSKMELKRGVSS 1437

Query: 1441 TLVEKFDMEGILGNLPEDNQLSLFDFFF 1468  
 TLVEK DEMGILGN+PEDNQLSLFDFFF  
 Sbjct: 1438 TLVEKQDEMGILGNMPEDNQLSLFDFFF 1465

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 753

- 25 A DNA sequence (GBSx0800) was identified in *S. agalactiae* <SEQ ID 2315> which encodes the amino acid sequence <SEQ ID 2316>. Analysis of this protein sequence reveals the following:

Possible site: 24  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1505 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 10179> which encodes amino acid sequence <SEQ ID 10180> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13207 GB:Z99111 similar to transcriptional regulator (MarX family) [Bacillus subtilis]  
 Identities = 49/124 (39%), Positives = 73/124 (58%)

Query: 18 VMKAFRTIDGKVSSEFKFELTPQFAVLQVLYAKGTWKIGSLIENMIATSGNMIVVIK 77  
 V +AF+++ KE PT+FAVL++LY +G K+ ++ +L SGN+T VI  
 Sbjct: 20 VPARAFKVSSEHSIRDSKEHGPNTEFAVLELYTRGPQKLQIGSRLLLVSGNVYID 79

Query: 78 NMEKKGVILRHSCPNCKRAFLVSL/TEGSEVIKKALPEHILKRVEDAFSVLTETEQLDLIN 137  
 +E+ G++R P DKR+ LT +G E+ K P H R+ AFS L+ EQ+ LI  
 Sbjct: 80 KLERNGFLVREQDPKDKRSVYAH/IDKGNEYLDKIYPIHALRTARAFSGLSPEDEQOLIV 139

Query: 138 LLKK 141  
 LLKK  
 Sbjct: 140 LLKK 143

- 50 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2317> which encodes the amino acid sequence <SEQ ID 2318>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

-855-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0537 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 80/145 (55%), Positives = 111/145 (76%), Gaps = 1/145 (0%)

Query: 2 GDEMGNF-KNSAVKSMVVRKAPRTIDGKVSSEFKFELTPTQFAVLVDLYAKGTWKIGS 60  
 G+M + KN+A+K+MVV RKA RT+D ++ FK+ +LT TQF+VL+VLY KG M+I  
 Sbjct: 8 GNQMSHLDKNLAKAMVVRKAQRTLDAPGADIFKIKADLTATQFVLEVIYTKGCMRINH 60

Query: 61 LIENMLATSGNMIVVIGKMEKGVLRHSCPNDRAFVLSLTTEGREVKKALPEHIKRV 120  
 LI++LATSGNMIVV+ NMB+ GW+ + DKRA++V+LT +G +I+ LP+H+ RV  
 Sbjct: 68 LIDSLATSGNMIVVILNMERNHGWSKCKDKRATVTVTLTRGTRLIAVLPHKVARV 120

Query: 121 EDASFVLTEQEQLINLLKKFKTL 145  
 E+AF+VLTE EQ LI LKKFK L  
 Sbjct: 128 EEAFAVLTEKQLCLIELLKKFKQL 152

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 754

A DNA sequence (GBSx0801) was identified in *S.agalactiae* <SEQ ID 2319> which encodes the amino acid sequence <SEQ ID 2320>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3742 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GPIAG05963 GB:AB004686 hypothetical protein [Pseudomonas aeruginosa]  
 Identities = 115/203 (56%), Positives = 143/203 (69%), Gaps = 7/203 (3%)

Query: 2 SFLEELKNRSIYALGRNTEVSDKEKIVEIKKAVRQSPSAPFNSQTSRVVILINDEVTWKF 61  
 +FL +KNR+LYAL+ VS EKIVE++KRAV SPSAPFNSQ+SRV+L E +FW  
 Sbjct: 4 AFLSSIKNRRTIYALDKQLPVSQEKIVELVKEAVSHSPSAPFNSQSSRVVILPQAEHQFV 63

Query: 62 DELVANDELVTMKVQCAPETAIAGTIRKILASPGASKGTVLFFEDQDVVKSILQBFVLYAD 121  
 + +A D E K+ P A A T+ KL SF A GTVLFFEDQ VV+ LQBQF LYAD  
 Sbjct: 64 N--IAKD--ELKKI--VPADAFAATETKINSFAAGSTVLFFEDQTVVRQLQBFQALYAD 117

Query: 122 NFFVWSBQSTGTSIVNTWTALSLAELGLGNNLQHYNPVIDASVQAVYGVPSMKLRGQLNF 181  
 NFFVWSBQ+G+A WTAL AE +G +LQHYNP++DA + +P SWKLR Q+ F  
 Sbjct: 118 NFFVWSBQASMAQFAVWTAL-ARHKVGASLQHYNPVLDVDAQTHKTWNLPESNKLRAQMPF 176

Query: 182 GSIEAETGEKSFNNDDRFKVIG 204  
 G+I A GEK F+ + +RFKV G  
 Sbjct: 177 GATAAFAGEKAPFAKSERFKVFG 199

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 755**

A DNA sequence (GBSx0802) was identified in *S. agalactiae* <SEQ ID 2321> which encodes the amino acid sequence <SEQ ID 2322>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2730 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB62846 GB:AL035475 hypothetical protein [Plasmodium falciparum]
(ver 2)
Identities = 112/529 (21%), Positives = 217/529 (40%), Gaps = 67/529 (12%)

Query: 3 NKGGKLLKNIEEFKTIQKRLTERGKFPYDTHVSTFEIKDENFIMERLKSSGLSMGKD-- 60
N K+ +K + ++ Q + E+ KF D H E + E P I E + + K
Sbjct: 1063 NVKIKNEKGAKN-DSLNQHEIIEKEKF--DLQH---ENRSEFIEEKKICIVDDKIKNNI 1116

Query: 61 --VDYMGVNGIPIYTKLISVKNFAPENNSKDSYSSNNINISDEKIKENDQKILDLIVKS 118
VD + P Y + L + + N + YS+ DK I +N++ ++ K
Sbjct: 1117 MNVDEKRRKSDHPSPYERVLKMEG-----SNKNEEGYSNT-----DKILANKKEIKGVNKKK 1166

Query: 119 GANNQNLTDSEKVIAPTKYIGETITNYDNEAYRARNVDTEYRASDLFSVTERKIAMCVGY 178
G N++ +E+K K + E + ++E D + F + ++C
Sbjct: 1167 GENDEKNEKNEKRENDKVNKIKDENEKNEKNEKIDENDNNNSYFYNNISDTTFELCTNS 1226

Query: 179 SVTAARAFNIMGIPSYVSGSKSPQGISHAAVRAYTYRSHWIIDITASTYWKNGNYKITYS 238
+ N + IPS ++ +GI + N S I+ KN N ++ YS
Sbjct: 1227 LIPINNKNISILIPS-----ENKGLIGSQKEEQNISPVKIDNNKKDKLCKNN-ESDYVS 1280

Query: 239 DFIKEYCIDGYD--VYDPAKTNFRFK-VKYMESNEAFENIWHNNGSKSML-----FIN 288
D ++ + +Y +N++ ++ ++ NE + + + N S++ L ++
Sbjct: 1281 DKQYSVLLNSIEKKIKYKCSNSKIRGIEKKKINEDYVDLGNINCSNPTLEFFLTLYKYLK 1340

Query: 289 TSAALKKKPKKDFPVPTKEKKNELIDKYKKLLSQIPENTQNGEKNIRDYLNKEYEYIL 348
S + ++ + V EK+K + K KKL +I N P + I + ++EY +
Sbjct: 1341 SSELINHEHQNDINNVYEKKIKKQAK-KKLNRKI--NVNIPNDGIEENMSSEYFVK 1397

Query: 349 KKN-----LPEHEHAE-----PKSLNLRSEFYQLKKER-----MKPSDNLKKEE 390
KK+N FE + ++ F N + L +E+ ++ +N K+ E
Sbjct: 1398 KNNNMMVKFETKRSKSLSSSEIFAVKNNKKRATNLMRSEBQFISSGLVKEGKNGRIE 1457

Query: 391 KPRENVSKERETPAENDFVSVTEKNNLIDKYKELLSKIPENTQNGEKNIRN--VLEKE 448
+ E +KE+ +N+F KNNL +++ L K EN G N +++
Sbjct: 1458 EKDEYIKKK-INKKNKF-----KNNLTQL--LPFKSARINISGSPNTEKIRHVKRT 1509

Query: 449 YEELQKDKLKFHEYTEFTKSLNINETFYSQLKGGEMKLSNPFGKGTIN 497
++ + + ++ K L E ++ E + ++++N EKGE N
Sbjct: 1510 KKKVLSNNFTLANNPSNLIKLLKLRMEKDKIKMDQKSKINKKNEKGRFN 1558

```

There is also homology to SEQ ID 598.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 756**

A DNA sequence (GBSx0803) was identified in *S. agalactiae* <SEQ ID 2323> which encodes the amino acid sequence <SEQ ID 2324>. Analysis of this protein sequence reveals the following:

-857-

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

5       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1243 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

10   No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 757

15   A DNA sequence (GBSx0804) was identified in *S.galactiae* <SEQ ID 2325> which encodes the amino acid sequence <SEQ ID 2326>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

20       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1057 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25   The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate  
           aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]  
   Identities = 78/192 (40%), Positives = 118/192 (60%), Gaps = 6/192 (3%)

30   Query: 14 KIVAVIRGNSCEEAFOAACACIKGGISAEIAYTNSKASQVIEQLVTQYTNQEQVWVG 73  
           KIVAVIR NS EEA + A A +GG+ IEI +T A VI++L + ++ ++GAG  
   Sbjct: 11 KIVAVLRANGVEEAKKALAVFEGGVHLIEITFTVFDADTVIKEL--SFLKKGKATIGAG 68

35   Query: 74 TVLDSSETARMAILAGAKTIVSPAPHLQTAKLCNKYALPYLPGCMTESEVYTALEAGCKRII 133  
           TV E R A+ +GA+FIIVP + + ++ C + Y+PG MT +E+ A++ G I+  
   Sbjct: 69 TVTVEQCKRKAVESGAEFIVSPLDDEISQFCKEKGVFTFPGVMTPELVKAMKLGHTIL 128

40   Query: 134 KIPFGTGLTGSFISSKAPLPQVQIMVTGGNLTNAKDWPLSGVTAIGGGEFKNLAALG 193  
           K-FPG +G F+ ++K F P V+ + TGGVNL N -WF +GV A+G+G K G  
   Sbjct: 129 KLPFGEVVGPQFVXAMKGGFFPNKVFVTGGVNLNDVCEWFKAGVLAVGVSSALVK----G 184

45   Query: 194 EFDKITEMAKQY 205  
           D++ E AK +  
   Sbjct: 185 TPDVREKAKAF 196

There is also homology to SEQ ID 1252.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 758

50   A DNA sequence (GBSx0805) was identified in *S.galactiae* <SEQ ID 2327> which encodes the amino acid sequence <SEQ ID 2328>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

-858-

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5                   bacterial cytoplasm --- Certainty=0.4213 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]  
     Identities = 94/329 (28%), Positives = 169/329 (50%), Gaps = 7/329 (2%)
- Query: 3   XILFFGRPLIRITPKENDYFADSIKLYFGSEVNTARALQGFQGTLLSALFNNPIG 62  
           K++ PGR ++R++P ++    + S + YG+K N A L G D ++ IPNNP+G
- 15 Sbjct: 2   KVVTGPEIMLRSLSPDHRKIPQTSFDVTVYGAEMVAAPLAQMGLOAYFVTKLPANNPLG 61
- Query: 63   NSFLQFLKAQGDITHSIQWVGSRVGLYPLBDSFACKRGSEVYDRHSSLDHFRINQIDFD 122  
           ++   L+ G+ T I G R+G+YFL + R +VYDR HS++ + + D++
- 20 Sbjct: 62   DAAAGHLRKFGVKTDYLARGGNRIGIYFLIGASQRPISKVYDRAHSALSEAKREDFWE 121
- Query: 123   QLFEGVSLFHFSGITISLDSIQRTITLLKKAKKREITISLDLNFPSKLLISPNKAKILF 182  
           ++ +G FHFSGIT L + + I LK A ++ +T+S DLN+R++L + + A+ +
- 25 Sbjct: 122   KILDGARWFHFSGITPFLGKELPLILELALKVANEKGVTSCLNLYRARLWTKEEAQKM 181
- Query: 183   SQFATFADICFG----IEFLMVDGQDTTFPNRDEATIEDVKERMISLNHFQDFQVIFHTK 238  
           F + D+   IE ++ S +    + Z +    + +F+ + T
- 30 Sbjct: 182   IPMEYVDVLLANBEDIEKVLGISVSGDLDTGKLNREAYAKIAEBVTRKNFPTKVGITL 241
- Query: 239   RLQDEWGRNHYQAYI-ANRQEFVTSKEITTVNQRIGSGDAFVAGALYQLLCHSDSKTV 297  
           R   N++ + N + F EI + R+G+GD+F   +Y L DS+
- 35 Sbjct: 242   RESISATVYVWSVMVFENGQPHFENKYEI--HIVDRVAGDGSFAGALYGLMGPDSQNK 299
- Query: 298   IDFAVASASLCKALEGDNMFPETVTVANNK 326  
           +FA R++ LK + GD + ++ + R+
- Sbjct: 300   AEFAAASCLKHTIPGDFFVLISIEIEKL 328

There is also homology to SEQ ID 1264.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 40 Example 759

A DNA sequence (GBSx0806) was identified in *S. agalactiae* <SEQ ID 2329> which encodes the amino acid sequence <SEQ ID 2330>. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence  
 45 INTEGRAL Likelihood = -0.22 Transmembrane 53 - 69 ( 53 - 70)

----- Final Results -----

- 50                   bacterial membrane --- Certainty=0.1086 (Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 55 >GP:AAD36157 GB:AE001768 sugar-phosphate isomerase [Thermotoga maritima]  
     Identities = 41/125 (32%), Positives = 61/125 (48%), Gaps = 10/125 (8%)
- Query: 1   MKIALINENSQASKNTIYKELKAVSDEKGFVFNQYMGYKEESQLTYVQNGLL/TAAIL 60  
           MKIA+ +++++   + ++K   KG EV ++G Y +E   Y + ++ +LL
- Sbjct: 1   MKIALASDHAAPE----LKEKVKNYLLKGKIEVDHGTYSSESDVDPDYAKK-VVQSILS 55



-859-

Query: 61 NSGAADFVITGCGTGIGAMLAQNSFPFGVCGFAADPVDAYLFSQVNGGNALSIPFAK3FG 120  
 N ADF I CQTG-G +A N +G+ P A L N N L L P G  
 Sbjct: 56 NE--ADFGILGCTGLGMSIAANKYRGIRALCLFPDMARLARSHNNNTLVLE---GRL 110

5 Query: 121 WGAEI 125  
 GAEL  
 Sbjct: 111 IGAEI 115

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2331> which encodes the amino acid sequence <SEQ ID 2332>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2599 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 159/212 (75%), Positives = 186/212 (87%)

Query: 1 NKIALINENSOAKNIIYKELKAVSDEKGPVFNHYGMVGBEESCLTYVQNGLLTAILL 60  
 NKIALINENSOAKNIIY L V+D+ G+VFNHYGMV E ESCLTYVQNGLL +ILL  
 Sbjct: 1 NKIALINENSOAKNIIYDALITVTDKHYQVFNHYGMVGBEESCLTYVQNGLLASILL 60

Query: 61 NSGAADFVITGCGTGIGAMLAQNSFPFGVCGFAADPVDAYLFSQVNGGNALSIPFAK3FG 120  
 + AADFV+TGCTG+GAMLA NSFPGV CGFA+P +AYLFSQ+NGGNALS+PFAK3FG  
 Sbjct: 61 TTGAADFVVITGCGTGVGAMLAQNSFPFGVCGFAADPVDAYLFSQVNGGNALSIPFAK3FG 120

Query: 121 WGAEINLRYLPERLFDKDGKGYPKERAVPEQFNARILSEIKQITVYRDLKSLVEIKDQDF 180  
 WGAEINL +FERLF + GGYPKERA+PEQFNARILS++K+TYRDL++K+IQDQF  
 Sbjct: 121 WGAEINLTLRPERLFAEPWGGYYPKERAIPBQFNARILSDKKITVYRDLATVVDKIDQDF 180

Query: 181 LKETISGEHPQEFYFFANCKQNIADYLSVLD 212  
 LKETISG HPQEFYFFAN + + YLKSVL+  
 Sbjct: 181 LKETISGAHPQEFYFFANAEPSSELVLYLSVLE 212

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 760

A DNA sequence (GBSx0807) was identified in *S.agalactiae* <SEQ ID 2333> which encodes the amino acid sequence <SEQ ID 2334>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.37 Transmembrane 10 - 26 ( 8 - 26)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 761

A DNA sequence (GBSx0808) was identified in *S. agalactiae* <SEQ ID 2335> which encodes the amino acid sequence <SEQ ID 2336>. This protein is predicted to be gluconate 5-dehydrogenase (fabG). Analysis of this protein sequence reveals the following:

```

5 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.1117 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:AAC77223 GB:AE000497 5-keto-D-gluconate 5-reductase [Escherichia
 coli K12]
 Identities = 116/260 (44%), Positives = 165/260 (62%), Gaps = 6/260 (2%)

 Query: 6 LKDNFSLBEGKVALITGASYGIGFSIATAPARAGATIVFNDIKQELVDKGISAYKKLGK 65
 + D FSL GK LITG++ GIGF +AT + GA I+ NDI E + + + GI+A
20 Sbjct: 1 MNDLFSLAGKNILITGSAQGIQFLATGLGKYGAQIIINDITAEARLAVKELKHQEGIQ 60

 Query: 66 HGYVCDVTDEEDGINBMVDKISQDVGVIDILVNNAGIIRKTPMLEMSAADPROVIDIDINA 125
 +VT + I+ V+ I +D+G ID+LVNNAGI +R P E ++ VI ++ A
25 Sbjct: 61 VAAPFNVTXKHEIDAAVEHIEKDIGFDVLVNNAGIQRHHPTTFPEQENVDVAVNQTA 120

 Query: 126 PFIIVSKAVLPQMIQKHGKIINICSMMSSELGRETVAAYAAAGKGLKMLTKNIASEYGSAN 185
 F+VS+AV M+++ GK+INICSM SELGR+T+ YAA+KG +KMLT+ + E N
30 Sbjct: 121 VFLVQSQAVTRHVMERKAGKVINICSMQSELGRDTITPYAASKGAVKMLTRCMVELARHN 180

 Query: 186 IQCNGIGPGYIATPQTAPLRERODGSRHFFDQFIIAKTPAARWGAEARDLGAIPAIFLSD 245
 IQ NGI PGY T T L E + F ++ +TPAARWG+ ++L A+FL+AS
35 Sbjct: 181 IQVNGIAGPYFKTENTKALVEDE-----AFTAWLCKRTPAARWGDPOELIGAAVFLSSK 234

 Query: 246 ASNFINGHILYVDGILLAYI 265
 AS+F+NGH+L+VDGG+L +
 Sbjct: 235 ASDFVNGHLLFVDGGMINAV 254

```

There is also homology to SEQ ID 1242:

```

40 Identities = 225/264 (85%), Positives = 246/264 (92%)

 Query: 6 LKDNFSLBEGKVALITGASYGIGFSIATAPARAGATIVFNDIKQELVDKGISAYKKLGK 65
 +++ FSL+GK+ALITGASYGIGF IA A+A+AGATIVFNDIKQELVDKG+++AY++LGI+A
45 Sbjct: 1 MRNMFSLQKALITGASYGIGFELAKAYAQAQATIVFNDIKQELVDKGLAAYRELGETA 60

 Query: 66 HGYVCDVTDEEDGINBMVDKISQDVGVIDILVNNAGIIRKTPMLEMSAADPROVIDIDINA 125
 HGYVCDVTDE GI +MV +I +VG IDILVNNAGI+RTPELM+ A DPROVIDIDINA
50 Sbjct: 61 HGYVCDVTDEAGIQCMVQSDIEVGAIDILVNNAGIIRKTPMLEMSAADPROVIDIDINA 120

 Query: 126 PFIIVSKAVLPQMIQKHGKIINICSMMSRIKRETVAAYAAAGKGLKMLTKNIASEYGSAN 185
 PFIIVSKAVLP MI KGHGKIINICSMMSRLGRETVAAYAAAGKGLKMLTKNIASE+G AN
55 Sbjct: 121 PFIIVSKAVLPQMIKANGHKIINICSMMSRLGRETVAAYAAAGKGLKMLTKNIASEPGEAN 180

 Query: 186 IQCNGIGPGYIATPQTAPLRERODGSRHFFDQFIIAKTPAARWGAEARDLGAIPAIFLSD 245
 IQCNGIGPGYIATPQTAPLRERQ DGRHFFDQFIIAKTPAARWG EDL FA+FLSD
60 Sbjct: 181 IQCNGIGPGYIATPQTAPLRERODGSRHFFDQFIIAKTPAARWGTTEDLAGPAVFLSD 240

 Query: 246 ASNFINGHILYVDGILLAYIGKQP 269
 ASNF+NGHILYVDGILLAYIGKQP
 Sbjct: 241 ASNFVNGHILYVDGILLAYIGKQP 264

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 762

A DNA sequence (GBSx0809) was identified in *S.agalactiae* <SEQ ID 2337> which encodes the amino acid sequence <SEQ ID 2338>. This protein is predicted to be mannose-specific phosphotransferase system component IIB. Analysis of this protein sequence reveals the following:

Possible site: 24  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0886 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD46485 GB:AF130465 mannose-specific phosphotransferase system  
component IIB [Streptococcus salivarius]  
Identities = 43/107 (40%), Positives = 61/107 (56%), Gaps = 3/107 (2%)

Query: 2 IKIIVAHGNFPDGLSSLELIAGHQEVVVGINFIAQMSSNDVRVALQREVIDFK---EI 58  
I III +HG F +GI S +I G QE V + F+ +D+ + F EI  
Sbjct: 3 IGIIAGSHOKFABGIIHSGSGMIFDQEKVQVTFMPSEGPDDLVAHFNDLIAQFDADDEI 62

Query: 59 LVLTLGLGTFPPNVSSALSVEYTDKKIKVLSGLNLSMLMEAVLGRIM 105  
LVL DL G+PFN +S ++ E D+KI +++GLNL ML++A R M  
Sbjct: 63 LVLADLWSGPPNQASRIAGENPDKIATITGLNLPHLQAYTERNM 109

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2339> which encodes the amino acid sequence <SEQ ID 2340>. Analysis of this protein sequence reveals the following:

Possible site: 41  
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF81086 GB:AF228498 AgaF [Escherichia coli]  
Identities = 48/127 (37%), Positives = 71/127 (55%), Gaps = 6/127 (4%)

Query: 1 MIAIIVMGHGFASGIUSAELIAGKQEKVTAIDFTTMTAADVDQLSKALIP---EER 57  
M++II+ GHG FASG+ A++ I G+Q + AID + A + QL A+ E+  
Sbjct: 1 MLSTILTGHGFPASGMRKAMKQIILGEGSQFIADIVPETSSTALLSQLEALIAQLDCEDG 60

Query: 58 TLVLCDLLSGTTPFKVAATLMSLENTTTCNVLSGLNLAIRASFAQTAAFSFDLVSGLI 117  
+ L DLLGGTTP+VA+TL P C V++G NL +L+E R+ + + V L  
Sbjct: 61 IVFLTLDLGGTTPFRVASTLAWKQKQ--CEVITGTNLQLLEMLVLEREGLSGEEFRVQAL- 117

Query: 118 TCSKEGI 124  
C G+  
Sbjct: 118 ECGHRL 124

An alignment of the GAS and GBS proteins is shown below.

Identities = 73/146 (50%), Positives = 94/146 (64%), Gaps = 3/146 (2%)

Query: 1 MIKIIIVAHGNFPDGLSSLELIAGHQEVVVGINFIAQMSSNDVRVALQREVIDFKIILV 60

-862-

```

MI II++ HG+F GI+S+LELLAG QR V I+F M++ DV+ L R +I +K LV
Sbjct: 1 MIAIIVAGHGHPAGSIVSALPELLAGKQKVTAIDFTTETAADVQQLSRALIPSETLV 60

Query: 61 LDTDLGGTFFPNVSSAISVEYTDKKIKVLSGINLSMLMEAVLSRIMFHEHVDIVKVTSS 120
L DDLGGTFF V++ L + VLSG:NL+ML+EA +R DDIV +IT S
Sbjct: 61 LCDLLGGTFFKVAATMBSSLPNTTQNVLSGINLAMLIEASPARQATASPDIVSGILITCS 120

Query: 121 HEGIVD*STCLATQTAEATPE--GGI 144
EGIVD* T L+ Q AT + GGI
Sbjct: 121 REGIVD*TKT-LSQQEDGATDDELGGI 145

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 763

A DNA sequence (GBSx0811) was identified in *S. agalactiae* <SEQ ID 2341> which encodes the amino acid sequence <SEQ ID 2342>. This protein is predicted to be unsaturated glucuronyl hydrolase. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.11 Transmembrane 172 - 188 (172 - 188)

----- Final Results -----
bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA05773 GB:AP001514 unsaturated glucuronyl hydrolase [Bacillus halodurans]
Identities = 156/370 (42%), Positives = 219/370 (59%), Gaps = 3/370 (0%)

Query: 30 EEAEIKGLKQLYINIDYFGEEYTPATFNNIYKVMNDTENTWNGFTWGLNLAYEYNQDKK 89
++A+ ++ NI F +P + Y++ +N EWNGFW+G LNL YEY D
Sbjct: 4 KQAMTDVAEKTLLNIKRPNRFRPHVSEDEGHEYLNNNNWNGFWSGILMLCYEYINDPA 63

Query: 90 LKNIHQNVLSFIANRINNRIALDHDHDLGFLYTPSCTAEYRINGDVGLEATIKAADKIME 149
+ A V SF R+ + LDHHD+GFLY+ S A++ I D +A + TI+AD LM+
Sbjct: 64 FRQAASTVRSFQQRMBQNLELDHHDHDLGFLYLSLSKQWIERDERAKQLTIRAADLMMK 123

Query: 150 RYQEKQGFQAWGELG-YKHEHYLLIIDCLNIQLLFPAYEQTGEKYRQVAVNHFYASAN 208
R++EK QAWG G R+I+DCL+N+ LLF+A E TG+ YR+ A+ H +
Sbjct: 124 RWREKIELFQAWGPEGLDSNGRIIVDCLMNLPLFWASEVTGNFDYERAAIHADKTRR 183

Query: 209 NVVRDSSASPHTFYFDPETGEPLKGVTRQGSYDRSSWARGWQAWGIYGLPSYRKMIDYQ 268
+VR D S +HTFYF+ ETGE L+G T QGY D S+W+RGQAW IYG ++YR + +
Sbjct: 184 PIVRGDSTYHTFYFQETGEALRGTHQGYEDGSTWSRGQAWIYGFALAYRYTCNEYR 243

Query: 269 IILFEGMTRYFLNRLPEDKVSYDMLIPTDGSQPRDTSATATAVQGIHEMLKYLPEVDPD 328
+ K YF+ LP D V+YWD RD:SA+A CGI E+L +L E DPD
Sbjct: 244 LETAKRITACYFIENLPADYVAYWDFNAPITPTDKRDSASATASGILELLSHQSTDP 303

Query: 329 KETYKYAMHTMLRSLIEQYSNNELIAGRPLLLHGVYSNHSKGVDBGNIGWDYFYLEALI 388
K ++ + + SL+E Y++ + G L+ G YS G D+ IWGDY+Y EAL+
Sbjct: 304 KAFFQCSVQRKMTSLVNTYASEKDAQG--LIRGGSYSVRIGHAPDYVWGDYFYFTEALN 361

Query: 389 RPYKWKELYW 398
R K YW
Sbjct: 362 RLEKLRNGYW 371

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2343> which encodes the amino acid sequence <SEQ ID 2344>. Analysis of this protein sequence reveals the following:

-863-

Possible site: 33  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.37 Transmembrane 173 - 189 ( 173 - 189)  
 ----- Final Results -----  
           bacterial membrane --- Certainty=0.1150(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 273/395 (69%), Positives = 336/395 (84%)

Query: 4 IKPVKVESIENPKRFLNSRLITKIEVEEATKALKQLYINIDYFGEYPTPATFNNIYK 63  
 +K + +E I+ P+RF L++ ++ +A++ ALKQ+ +N+DYF E++PTPAT +N Y +  
 Sbjct: 5 LKTIALEPIKQPERFTKEDFLSQEDITQALOLAKQLVRLNMDYFKEDFFTPATKQYAI 64

Query: 64 NDNTETWINGFWTGCLWLAYEYNQKKLKNIAHQNVLSPFNKRINRIALDHHDLGFLYTPS 123  
 NDNTETWLN FWTGCLWLAYEY+ D +K +A N LSP+R+ I LDHDLGFLYTPS  
 Sbjct: 65 NDNTETWLNFWTGCLWLAYEYSGDDAKALQAANDLSFLDRVTRDIELDHDLGFLYTPS 124

Query: 124 CTASYRINGDVKALEATIKAADKIMERYQKGGFTQAWGSELGYKEHYRLIIDCLNIIQLL 183  
 C AE+++ ++ EA +KAADKL++RYQ+KGGFTQAWGELG KE YRLIIDCLNIIQLL  
 Sbjct: 125 CMAWKLKLTPESEAAALKAADKI+VQRYQKGGFTQAWGELGKEDYRLIIDCLNIIQLL 184

Query: 184 FFAYEQTGDEKRYQAVNHFYASANNVVRDSSAFHTFYFDPETGEPKLVTRTQGYSDS 243  
 FFA ++TGD +YR +A+NHFYASAN+V+RCD+SA+HTFYFDPETG+P+KSVTRTQGYSD+S  
 Sbjct: 185 FFASQGTGDNRYRDMALNHFYASANHVIDOASAYHTFYFDPETGDFKGVTRTQGYSDS 244

Query: 244 SWARGQAWGIYGIPLSYRMRKDYQIILFRGMTNVLNRLPSDKVSYMDLIPTDGSQQR 303  
 +WARGQAWGIYGIPL+YR +K+ + I LFRGMT+YFLNRLP+D+VSYMDLIF DGS Q R  
 Sbjct: 245 AWARGQAWGIYGIPLTYRFLKEPELILQFGRMTHYFLNRLPKDQVSYMDLIPDGSQQR 304

Query: 304 DTSATATAVCGIHMLKYLEPVDPOKETIKYAMHTMLRLSLIQSYNNELIAGRPILLAGV 363  
 D+SATATAVCGIHMLK LP+ DPK+TY+ AMH+MLR+LI+ Y+N +L G PLLLHGV  
 Sbjct: 305 DTSATATAVCGIHMLKTLDPDHPDKTYEAMHSMRLALIKDYANKDLKPGAPILLHGV 364

Query: 364 YSNHSGRGVDEGNIWGDYYLLEALIRFYKDWELW 398  
 YSNHSGRGVDEGNIWGDYYLLEAL+RPHYCW YW  
 Sbjct: 365 YSNHSGRGVDEGNIWGDYYLLEALLRIFYKDWNPYW 399

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 764

A DNA sequence (GBSx0812) was identified in *S. agalactiae* <SEQ ID 2345> which encodes the amino acid sequence <SEQ ID 2346>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3035(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:RAC44679 GB:U65015 PTS permease for mannose subunit IIIMan C  
 terminal domain [Vibrio furnissii]  
 Identities = 63/125 (50%), Positives = 89/125 (70%), Gaps = 1/125 (0%)

Query: 5 PNIVMTRVDERLIHQG-GQLWVKPLSCNTFVIVANDDVSVDHLQQTLMKTVVPESIALRFP 63  
 PNIV++R+DERLIHQG G WV F N V+VAND+V+ D +QQ LM+ V+ + IA+RF+

-864-

Sbjct: 2 PNIVLSRIDERLWHQVGVQVQWGVADANIVVVANDEVAADTIQQNLMEWVLADGLAIRFW 61

Query: 64 DIQKVIDIIHKANPACTIFIIVKDLKDVYELVAGGVPIKEINIGNIHKGEGKQVRSRIF 123  
+OK ID IHKA+ Q I ++ K D RLW GGVPI IN+GN+H +GK Q+S++

Sbjct: 62 TVQKTTIDTIHKASDRGRILLVCKTTPHDFRLLVBSGVPIAANVGMHVIIDGKTQISKTVS 121

Query: 124 LGMKD 128  
+ +D

Sbjct: 122 VDAED 126

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2347> which encodes the amino acid sequence <SEQ ID 2348>. Analysis of this protein sequence reveals the following:

Possible site: 58  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2511 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA084216 GB:AB019619 unsaturated glucuronyl hydrolase [Bacillus  
sp. GLL]  
Identities = 161/369 (43%), Positives = 220/369 (58%), Gaps = 1/369 (0%)

Query: 32 QALDIALKQVRLNMEDYFKEDPPTPAIKNQYAIMDNTMTNAFWTGCMLAYEYSGDGAI 91  
QA+ ALK N+ F + PP + N+Y + DNT+MT+ FW+G LML YEY+GD+  
Sbjct: 4 QAIGDALGITARNLKKFGDRPFHVSDOSNKYVLNDNTDWDGFWGILMLCYEYTDDEQY 63

Query: 92 KALAQANDLSLDRVIRDIOLDDHIDGLFPLTSPCAEWKLLKTPSPSRALKAGADKLQVR 151  
+ A SF +R+ R LDHHD+GFLY+ S A+W + K +R+ AL AAD L++R  
Sbjct: 64 REGAVPTVASFRRLDRFENLDHHDIGFLYSLSAKAQWIVEKDSARKLALDAADVLMPR 123

Query: 152 YQKGGFICAWGELGKKEDY-RLI IDCLLNIQLLPASCQETGDNRYEDMAINHYFASANH 210  
++ G IQAWG G E+ R+IIDCLL+ LL +A ++TGD YR +A H S  
Sbjct: 124 WRADAGI IQAWGPKZDPENGR I I IDCLLNLPLLWAGECTYDPEYRVAEHALKSRFP 183

Query: 211 VIRDDASAVHTFFDPETGDPVKGVTRQGVSDSASARGQAWGIYGIPLTVRFLEPELI 270  
++R D S+YHTFYHDE G+ ++G T CG +D S W RQAWGIY G L R+L  
Sbjct: 184 LVRGDDSYHTFFDPENGINAIRGGTHCGNTDGTSTWTRQAWGIYGFALNSRYLGADLL 243

Query: 271 QLFKRMTHYFLNRLPKDQVSYNDLIFDGGSEQRDSSATAIACVGIHEMLKTLDPHDPDK 330  
+ K M +FL R+P+D V YND RDSSA+AI CG+ E+ L +DP++  
Sbjct: 244 ETAKRMARHFLARVFEDGVVYNDPEVFPQEPSSYRDSAGAITACOLLEIASQLDESDPER 303

Query: 331 KTYEAMHSMILRALIKUYANKDLKPGAFLLHGVYSMHSGKGVDEGNINWIDYYYLEALLR 390  
++ A + +AL YA +D + GY G D+ IWGDYYYLEALLR  
Sbjct: 304 QRFIDAAKTTVTALRDGYAERDDGEAEGRIRRGSYHVHSGISPDYITWIDYYYLEALLR 363

Query: 391 FYKDNPYW 399  
+ YW

Sbjct: 364 LERGVTYW 372

An alignment of the GAS and GBS proteins is shown below.

Identities = 112/160 (70%), Positives = 132/160 (82%), Gaps = 1/160 (0%)

Query: 5 PNIVNTRVDERLTHQCGQLWVKFPLSCNTVIVANDVSKDHLQQTMTKTVVHKSIALRFFD 64  
PNI+MTRVDERLTHQCGQLWVKFPL+CNVTIVAND VS+D +QQ+LMKTV+P SLA+RFF  
Sbjct: 4 PNI+MTRVDERLTHQCGQLWVKFLACNTVIVANDAVSEKIQQLSMKTVIPSSIALRFFS 63

Query: 65 IQKVIDIIHKANPACTIFIIVKDLKDVYELVAGGVPIKEINIGNIHKGEGKQVRSRIFL 124  
IQKVIDIIHK+PAQ+IFI+VKDL+D LV GGVPI EINIENIH + K +++ I L  
Sbjct: 64 IQKVIDIIHKASPAQSIFIVKDLQAKLLVBGGVPIEINIGNIHKTKDKVAITQFISL 123

Query: 125 GMDKEIIRKLNGEYHIAFWTKITPTGNDGAVZVNIIDYI 164  
 G DK IR L ++H+ FNTKITP GN A +V+IIDYI  
 Sbjct: 124 GETDKSAIRCLAHDEHVVFNTKITPAGN-SASVDIIDYI 162

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 765

- A DNA sequence (GBSx0813) was identified in *S. agalactiae* <SEQ ID 2349> which encodes the amino acid sequence <SEQ ID 2350>. This protein is predicted to be AgaW (agaC). Analysis of this protein  
 10 sequence reveals the following:

Possible site: 25  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -6.95 Transmembrane 251 - 267 ( 244 - 269)  
 INTEGRAL Likelihood = -4.30 Transmembrane 213 - 229 ( 208 - 230)  
 15 INTEGRAL Likelihood = -2.71 Transmembrane 149 - 165 ( 148 - 165)  
 INTEGRAL Likelihood = -1.81 Transmembrane 31 - 47 ( 31 - 49)  
 INTEGRAL Likelihood = -1.49 Transmembrane 173 - 189 ( 173 - 189)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3781(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:AAF81084 GB:AF228498 AgaW [Escherichia coli]  
 Identities = 93/295 (31%), Positives = 140/295 (46%), Gaps = 48/295 (16%)  
 Query: 1 MDISILQAVLGLNTAFCSQMLLGL-YNRCIVLSLGVGVILGDIQTALVAGVISELAY 59  
 M+IS+LQA +G+ M GL + +R +VL VG++LGD+ T + G EL +  
 30 Sbjct: 1 MEISILQAFALGIITAFGLDMFNGLTHMRFPVVLGPIVGLVGLDHTGILRGSTLELVN 60  
 Query: 60 MGFGVGAGSTVPPNPPIGPIGFTLMAITTAGTKKITPEAALALSTPIAVGIQFLQATY 119  
 MG AG PPN I I GT AITT + P+ A+ ++ P AV +Q T +  
 35 Sbjct: 61 MGLAPLAGAQ-PPNVIIGTIVGTAFATTTG-----VKPDVAVGVAVPFAVAVQMGITFLF 114  
 Query: 120 TAFAGAPETAKK-----ALCAGNFRGFKIAANGT-IWAFAGLGFGLGLVGLALSTCYL 170  
 + +G + AL A N+ N + AF + FG A +T+  
 Sbjct: 115 SVMGVMSCARMPTPTILAAINACNYLALLALGNFYFLCAFLPIYFG-----AEHTAKTI 169  
 40 Query: 171 TDLFALIPFVLNGLTAGKMLPAIGFAMILSVMAKKELIPYLLGYVLAVYFGLPVLT 230  
 D+ +P L++GL +AG ++PAIGFA++L M K IPY +LG+V A + LPVL  
 Sbjct: 170 IDV---LPQRLIDGLGVAGSIMEPAIGFAVLLKIMKNNVPIPYTILGFVAAWLKLPVL-- 224  
 Query: 231 TANGDGLVTSVATNSVLGVPTIGVAIITATIFALLDIPKPAAPTKEKTEGDNQD 285  
 +A A AL+D+ RK PT+ + + +D  
 45 Sbjct: 225 -----ALACPALAMALIDLKRSPEPTQPAQAQKEEFD 257

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2351> which encodes the amino acid sequence <SEQ ID 2352>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 52  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -6.37 Transmembrane 220 - 236 ( 214 - 241)  
 INTEGRAL Likelihood = -5.10 Transmembrane 146 - 162 ( 144 - 165)  
 55 INTEGRAL Likelihood = -1.59 Transmembrane 184 - 200 ( 184 - 202)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3548(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```
>GP:AA044680 GB:U65015 PTS permease for mannose subunit IIPMan
(Vibrio furnissii)
Identities = 86/255 (33%), Positives = 137/255 (53%), Gaps = 11/255 (4%)

Query: 1 MDINLLQALLIGLWTAFCFSGMLLGI-YTNRCILSRGVGIIQLDPTALSMGALISELAY 59
M+I L QAL++GL + G+ + +R ++L VG+ILGEL T + +G EL +
Sbjct: 1 MEIGLFLQALMLGLLAPLAGLGLDFNGLTIFHRPVLVIGPLVGLILGELHPTGLVGGTLELIW 60

Query: 60 MGFGVGAGGTVPPNPIGPGIRGTILMAITSAGKVTPEAALALSTPIAVIQFLQTFAYTAF 119
MG AG PPN I I GT AIT+ V P A+ ++ P AVA+Q T ++A
Sbjct: 61 MGLAPLAGAQ-PPNVIIIGTIVGTTFAITT--NVEWVAVGVAVFPAVAVQMGITLLFSAM 117

Query: 120 AGAPETAKKQLQKGNIRGFK--FAANGTIWAFIPLGLIGLGLSMDTLHLHVDYIPP 176
+ + + + RG + + A + + F F+ L + L D +V +P
Sbjct: 118 SAVMSKDEYAKNADTRGIERVNFALAVLSFYFLCAFLPIY--LGADHAGMVAALFK 175

Query: 177 VLLNGLTVAGKMLPAIGFAMILSVMAKKELIPVILGIVCAAYLQIPTIGIATIGIIFAL 236
L++GL VAG ++PAIGFA+++ +M K IP+ ++G+V AA+LQ+P + I A-
Sbjct: 176 ALIDGLGVAGGIMPAGFAVIMKIMMKVATPYFILGFVAAMWLQPLAIRCAATAGI 235

Query: 237 NEFTNK--PKQNDAT 249
+? K P V+A+
Sbjct: 236 IDFMKSEPTPVNAS 250
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 203/288 (70%), Positives = 225/288 (77%), Gaps = 28/288 (9%)

Query: 1 MDISILQAVLIGLWTAFCFSGMLLGIYTNRCIVLSIGVGVLIEDIQTALAVGAISELAYN 60
MDI++LQ++LIGLWTAFCFSGMLLGIYTNRCI+LS GVG+ILSD+ TAL++GAISELAYN
Sbjct: 1 MDINLLQALLIGLWTAFCFSGMLLGIYTNRCILSRGVGIIQLDPTALSMGALISELAYN 60

Query: 61 GFGVGAGGTVPPNPIGPGIRGTILMAITAGTKITPEAALALSTPIAVIQFLQTFATYT 120
GFGVGAGGTVPPNPIGPGIRGTILMAIT+AG K+TPEAALALSTPIAV IQFLT YT
Sbjct: 61 GFGVGAGGTVPPNPIGPGIRGTILMAITSAG---KVTPEAALALSTPIAVIQFLQTFATYT 117

Query: 121 AFAGAPETAKKALQAGNFRGFKIAANGTIWAFAGLGGFGLGVLGALSTQTLTDLFALIPPV 180
AFAGAPETAKK LQ GN RGFK AANGTIWAF +G GLG+LGALS TL L IPFV
Sbjct: 118 AFAGAPETAKKQLQKGNIRGFKFAANGTIWAFIPLGLIGLGLSMDTLHLHVDYIPPV 177

Query: 181 LNLGTLVAGKMLPAIGFAMILSVMAKKELIPYILGIVYLAAYFGLVPLVTPIANGDVLTS 240
LNLGTL+AGKMLPAIGFAMILSVMAKKELIP+L+GTV A Y
Sbjct: 178 LNLGTLVAGKMLPAIGFAMILSVMAKKELIPFVLIGTVCAAY----- 219

Query: 241 VATNSVLGVPTIGVVAIIATIFALLDIFRKEAAPTKEKTSQDNQDDWI 288
L +PTIG+ALI IFAL + + KF T +G QDDWI
Sbjct: 220 -----LQIPTIGIATIGIIFALNEPNKF-KQVDATTVQSGQDDWI 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 766

A DNA sequence (GBSx0814) was identified in *S. agalactiae* <SEQ ID 2353> which encodes the amino acid sequence <SEQ ID 2354>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2442 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```



The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 767

A DNA sequence (GBSx0815) was identified in *S.agalactiae* <SEQ ID 2355> which encodes the amino acid sequence <SEQ ID 2356>. This protein is predicted to be PTS permease for mannose subunit IIBMan.

Analysis of this protein sequence reveals the following:

```

10 Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.28 Transmembrane 278 - 294 (272 - 294)
 INTEGRAL Likelihood = -3.45 Transmembrane 155 - 171 (155 - 174)
 INTEGRAL Likelihood = -1.59 Transmembrane 250 - 266 (250 - 267)
15 ----- Final Results -----
 bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
20
```

A related GBS nucleic acid sequence <SEQ ID 8657> which encodes amino acid sequence <SEQ ID 8658> was also identified. Analysis of this protein sequence reveals the following:

```

25 Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -9.70
 GvH: Signal Score (-7.5): -6.12
 Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 3 value: -8.28 threshold: 0.0
 INTEGRAL Likelihood = -8.28 Transmembrane 254 - 270 (248 - 270)
30 INTEGRAL Likelihood = -3.45 Transmembrane 131 - 147 (131 - 150)
 INTEGRAL Likelihood = -1.59 Transmembrane 226 - 242 (226 - 243)
 PERIPHERAL Likelihood = 0.37 175
 modified ALOM score: 2.16
35 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA57943 GB:U18997 ORF_c290; Geneplot suggests frameshift
 linking to o267, not found [Escherichia coli]
45 Identities = 101/278 (36%), Positives = 164/278 (58%), Gaps = 6/278 (2%)
Query: 17 LRQKETTKTGSSKILAKSDYTKTALRAPYLQNGPHYSNYQCLGYANVLYPALKKYYGDK 76
 ++ K+ T GS+ ++K D T+ R+ LQ FHY Q G+ + P LKK Y DKK
Sbjct: 19 VNMKKRTTMSSE-ISKDITRLGFRSSLLQASPHYERMQGGFTWMLPILKKIYKDD 77
50 Query: 77 KALAGALEENVEFYNTNPHPLPFTVSLHMLDLNERPHEETRIGIKMALACPLAGIGDLS 136
 L+ A+++N+EF NT+P+ + P+ L ++M + + I+G+K+AL GP+AGIGD+
Sbjct: 78 PGLSAAMKINLEFINTHPNVGLFLMGLLISMEEKGKRDITIKGLKVALGPIAGIGDAIF 137
55 Query: 137 QCFCLAPLESTIAASLADTGLVWGLFFVAMNTILTGIKLVTCMYGYRLGTSFIDKLSQ 196
 F L P+ + I + S A+ G ++GPIILF A+ ++ +++ GY +G IDK+ E
Sbjct: 138 WFTLLPINDAGICSSPAGSQNLGLPILFF-AVYLLIPFLRVGNTHVGVSVVKAIDKVRN 196

```

- Query: 197 MSVISRAANIVGVTVISSLAATQVKLTIPYTPAPEKVTSTTQKIVTVQGNLKDIAAPLLE 256  
+I+R+A I+G+TVI L A+ V ++ +FA T + Q DK+ P +LP  
Sbjct: 197 SQMIARSATILGTVIGGLIASYVHINVVTSFA----IDNTHSVALQDFDPKVPNILE 252
- 5 Query: 257 ALTYTFMFLYLIKNNKNTTYKLVILTVIIGILGSMGLIL 294  
YT LW+Y ++ KK L+ +T ++ I+ S GIL  
Sbjct: 253 MAYTLLMYYPFLRVKKAHPVLLIGVTVFLSIVCSAPGIL 290
- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2357> which encodes the amino acid sequence <SEQ ID 2358>. Analysis of this protein sequence reveals the following:
- Possible site: 45  
>>> Seems to have no N-terminal signal sequence
- 15 INTEGRAL Likelihood = -8.49 Transmembrane 276 - 292 { 270 - 292 }  
INTEGRAL Likelihood = -7.01 Transmembrane 151 - 167 { 149 - 176 }  
INTEGRAL Likelihood = -3.03 Transmembrane 202 - 218 { 202 - 220 }  
INTEGRAL Likelihood = -2.13 Transmembrane 249 - 265 { 248 - 265 }
- 20 ----- Final Results -----  
bacterial membrane --- Certainty=0.4397 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- The protein has homology with the following sequences in the databases:
- 25 >GF:AA57943 GB:U18997 ORF\_c290; Geneplot suggests frameshift  
linking to c267, not found [Escherichia coli]  
Identities = 104/285 (36%), Positives = 162/285 (56%), Gaps = 7/285 (2%)
- 30 Query: 8 NKSMAQLSKENANVTGSKNLTIDYLTALRAFFLQNGFNYNQGGIGYANVTIPALGX 67  
N+E + + + + +KHD + R+ LQ FNY Q G+ + P LKX  
Sbjct: 13 NRSPLPVMMKKRTTAMGSEISKDITLGRSSLLQASFNTERMGAGGTWAMLPILKLI 72
- Query: 68 FGNQKGLYQALEDNCEFYNTNPHFLPFTISLHVLNENRPEETRNKIMALMGFLAGI 127  
+ +DK GL A+DN EF NT+P+ + P+ L + M E + + +K+AL GP+AGI  
35 Sbjct: 73 YKDKXPLSAAKONLEFINTENPLVGLMGLISMBEKGENDRTIIGLKVLPPIAGI 132
- Query: 128 GDSLSQFCIAPLFPSTIAASLASDGLVGLPILFFLAMHILTAIKGSGLYGVKGTSPID 187  
GD++ F L P+ + I +S AS G +LGPILFF A+ + + + +G GY VG ID  
40 Sbjct: 133 GDAIFWFTLLPIMAGICSSFSASQNLGLPILFF-AVTLILFFLRVGTWGVYGVKVID 191
- Query: 188 KLSGEMAVVSRMANIVGVTVIAGLAATSKVITVFITPAAGKVDAAATQKVFITQMLDK 247  
K+ E + + +R A I+G+TVI GL A+ V I V +FA + Q F DK  
Sbjct: 192 KVRENSQMIARSATILGTVIGGLIASYVHINVVTSFAIDNTHSVALQDF-----FDK 245
- 45 Query: 248 IAPALPALPFTLLMYLILKNNKNTTYKLVILTVIIGILGSMGLIL 292  
+ P +LP +TLLMYV ++ KK L+ +T ++ ++ S GIL  
Sbjct: 246 VFPNILEMAYTLLMYPLRVKKAHPVLLIGVTVFLSIVCSAPGIL 290

An alignment of the GAS and GBS proteins is shown below.

- 50 Identities = 224/288 (77%), Positives = 255/288 (87%), Gaps = 4/288 (1%)
- Query: 12 HLLKGLRQ--KETTQWIGSKKLAKSDYTKTALRAFFLQNGFNYSNQGIGYANVTIPALX 69  
+L K + +Q KE NMGS KL K DY KTAIRAF+LQNGFN+NYQG+GYANVTIPALX  
55 Sbjct: 6 NLSKMAQLSKENANVTGSKNLTIDYLTALRAFFLQNGFNYNQGGIGYANVTIPALX 65
- Query: 70 KYGDKKALAGALENVRFYNTNPHFLPFTISLHVLNENRPEETRNKIMALMGFLA 129  
K++G+DKK L ALE+N EFYNTNPHFLPFTISLHVLNENRPEETRNKIMALMGFLA  
Sbjct: 66 KHPGDKKGLYQALEDNCEFYNTNPHFLPFTISLHVLNENRPEETRNKIMALMGFLA 125
- 60 Query: 130 GIGDLSQFCIAPLFPSTIAASLASDGLVGLPILFFLAMHILTAIKGSGLYGVKGTSP 189  
GIGDLSQFCIAPLFPSTIAASLASDGLVGLPILFF+AMN ILL IK+ +G+YGY+GTSP  
Sbjct: 126 GIGDLSQFCIAPLFPSTIAASLASDGLVGLPILFFLAMHILTAIKGSGLYGVKGTSP 185
- Query: 190 IDKLSQMSVISRAANIVGVTVISSLAATQVKLTIPYTPAPEKVTSTTQKIVTVQGNL 247

-869-

IDKLSEQM+V+SR ANIUGVTI+ LAAT VK+T+P TFA KV +T QK VT+QGM L  
 Sbjct: 186 IDKLSEQM+V+SRMANIUGVTIAGLAATS VKITVPITFAAGKVDAANTAKFVTIQGM L 245

Query: 248 DKIAFALLPALVTFLMEFLYLNKKKWTYYKLVLTVIIGIGSWLGLIA 295  
 DKIAFALLPAL+T IM+YLNKKKWTYYKLVLTVIIG+GWSLGLIA  
 Sbjct: 246 DKIAFALLPALFTLIMYLNKKKWTYYKLVLTVIIGIGSWLGLIA 293

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 768

A DNA sequence (GBSx0816) was identified in *S. agalactiae* <SEQ ID 2359> which encodes the amino acid sequence <SEQ ID 2360>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.37 Transmembrane 135 - 151 ( 135 - 151)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB01924 GB: Z79691 OrfA [Streptococcus pneumoniae]  
 Identities = 76/206 (36%), Positives = 124/206 (59%), Gaps = 1/206 (0%)

Query: 428 SWITNSYPKCDYVQGLTSKDRYHLVEQQLHWQRASDIYHKRWLTLRQAITLVIDKVSFCP 487  
 SW Y YP +C ++ H+EG Y HGR +L L + + L+D + C  
 Sbjct: 2 SWIEYYPHSLFCHHKEREGRMYIEGAYWSAEPDLPYLHKRKLIMLVEDVWLLVDDIRQ 61

Query: 488 GRVLTNQVILLDDQIVYENGFPVNDLWVSTPTFNLEDCLISKRYNQLTSHSLKLVKKIKFV 547  
 G+H +Q+ILD V Y++G +N L+L S F+LED +IS +YN+L S KL K+ F  
 Sbjct: 62 GQHEALTCFILDKDVITYCGKINQLRIWSEVDFLEDITISPKYNELERSKSLTKRQFFE 121

Query: 548 DEVMQDTYLIVDRNCQVQYVPLVQTNSHIGLNSLAFDIRSQDFHYLIGVLMDDIIPGDKL 607  
 ++++DVT+I + ++ + QT+ +E+ N++AF++++ + LI +L +DI G+KL  
 Sbjct: 122 NQMLDYTTIAHSPETIRHSVYQTD-REVENALAFVNDETDKLILLSEDIRVGEKL 180

Query: 608 YLACQIKCKGKGVIVYDKRNGKMSRLK 633  
 L+ G K +GK +VYDK N +M RL+  
 Sbjct: 181 CLVDGTWGRGKCLVYDKINERIRLQ 206

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2361> which encodes the amino acid sequence <SEQ ID 2362>. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.55 Transmembrane 477 - 493 ( 477 - 493)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2020 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB01924 GB: Z79691 OrfA [Streptococcus pneumoniae]  
 Identities = 75/207 (36%), Positives = 125/207 (60%), Gaps = 2/207 (0%)

Query: 434 SWAYLSYPKENYCHLQNGHIVYFIRGSVYQTPSDRNNYQHDRQIILPGIFLIIDTIQA 493  
 SW Y YP S +CH ++ ++IBG+Y + D Y H R+IL+L ++L+D I+  
 Sbjct: 2 SWIEYYPHSLFCHHKEREGRMYIEGAYWSAEPDLP-YLHKRKLIMLVEDVWLLVDDIRQ 60

Query: 494 QGNHCLVSGQFILNHLVDKTDHLSDLRLISDCPFTTETILSKKYNQYLTSKHLKRRKF 553  
 QG H ++QFILD + + ++ LRL S+ F +E+TI+S KYN+ S KL KR+ F  
 Sbjct: 61 QGHEALITQFILDKVTYQDGKINQLRLMSVDLFDLTITISPKYNELRSSKILTRQFF 120

Query: 554 KDEGCTSTLLVDDTKVTPLTQTKGRNPIETALSHLKGKQFDYSICVLQDLKGEK 613  
 +++ T++ + ++ + QT R +E AL++ +K + D I +L ED+ GRK  
 Sbjct: 121 ENQMLDYTTIAHSEFIIIRHSVYQTDRE-VENALAFEVKNDETDLKILLISSEDIRUGEK 179

Query: 614 LVLNLSHKIRGKVVVNHITNRIIRLK 640  
 L L++ K+RGK +V + I +IRL+  
 Sbjct: 180 LCLVDGTFMRGKCLVDYKINRMIRLQ 206

An alignment of the GAS and GBS proteins is shown below.

Identities = 282/631 (44%), Positives = 414/631 (64%), Gaps = 2/631 (0%)

Query: 6 YNFKFD- FDEFPQCKYIKTYQSNAYQEMKASVNLAMNTFFVFNMDMMEPCSKAYCLDPL 64  
 + +FK+ + +FC+ YQ+++Y + K +L++ NTF+ F DNMDMEPC Y LDP+  
 Sbjct: 11 FARFKETVNPDFCFNYLLDYQTDSDYADQKRIADLLNTNPLFEDNMDMEPCPIYTHLOPI 70

Query: 65 ENDKPVTDDPENLQMLNRQTYLKFPLAVYIVEGDGSVLRQMKYPMYHNDQPTLKPEGA 124  
 M + V DDEPN +MLNRQTYL K +VY+VE D+ YL K P+ +NI+ L P+G  
 Sbjct: 71 TWQEAVIDDPENFMNLRQTYLQKLLVYLVERDERYLLTAKGFIAMWIESAIPDLPGEL 130

Query: 125 VSRITDGIKRMNSMLKVLPLDYFGLITETKKIKLLTSLREQITYMRDYREKDGSLNMG 184  
 +RT+DTGIRK +W+K LI+L+ F +T+ ++ +L SL +Q+ ++ Y +K SLNMG  
 Sbjct: 131 ATRITDGIKRCFANVKLIYVNLNPLNATLKQESSLILASLEKQLQFLHANYLDRKYSLSNMG 190

Query: 185 ILQTTAILACLYYTEDNLNLPETQSFARREILLQIKLQILDDGSQYBQSINHYHVEVLKSL 244  
 ILQTTAIL Y+ +L++ +FA +EL QI LQIL+DGSQ+BQS MYHVEVLK+L  
 Sbjct: 191 ILQTTAILADAYPGSDLDIAATAFARKELTQQIALQILEDGSQFQSTNHYHVEVLKSL 250

Query: 245 MELVILAPKYPLPLEETIERQVLYLIAMTGPDYCOLAGDSVDITDRILTLATLVKSS 304  
 +EL L P Y L T+ M YL+ MTGPD+ Q+ +GDSVDITDRILTLA +L+  
 Sbjct: 251 LELTALVPDYLQRLPTLAMS DYLLKMTOPDHKQIPLDGSVDITDRILTLAATILEEP 310

Query: 305 KTKSPSFNVNLETLLLPKPSIYLFEEIPRATIGESAYLPDPGSHVCLRDDRYILFFKN 364  
 K+ +F +++++LL G+ ++ FE++D T+ A+ F SGH+ + + Y+FFKN  
 Sbjct: 311 HLKAAAFPTLIDSLLLLGKGVHTPEQLPVQTLPTFAHNFPEHSGITINQENYILFFKN 370

Query: 365 GPPOSHTSDNNVCLYDKKKPIPIDAGRYTYKEQLRVDKRSSTSHSTCTDLOGQLEM 424  
 GP GS+HTHSD NS+CLY K +P+P DAGRYTYKEE LRY K + HST L+ Q E  
 Sbjct: 371 GPIGSSHSDQNSLCLYKQPLFCDAGRYTYKEEPLRYALKSRASHSTAFLESQLEQ 430

Query: 425 IKDSWYNSYPKCYQLTSKDRYHLVGGQLHVORAS-DIYHKRWLLTLPQATILVICK 483  
 I SW Y SYPK +YC L + +EG Q+ + Y H R +L LP I L+ID  
 Sbjct: 431 IDSWAYLSYPKSNYCHLRQNGHVFYIEGSCYQFSDRNMYQHRQLILPPGFILIIDT 490

Query: 484 VSCPGHVLNMQYLIDQGVYENGAVNDLKLVSPTTFNLEDCLISKRYNQLTESHKLKVK 543  
 + G H L+ Q+ILD+ + + +DL+L+S F +E+ +SK+YNO SHKL+K+  
 Sbjct: 491 IQAQGNHCLVSGQFILNHLVDKTDHLSDLRLISDCPFTTETILSKKYNQYLTSKHLKRRKF 550

Query: 544 IKFVDEVMYTLIVDRNQCKYKVLVQTNHSHKELNS IAFDIRESDPHYLIGVIMDDIIF 603  
 F D+ TL+V + +V + +OT + +++++ + F Y I VL +D+I  
 Sbjct: 551 KPFKDKGCTSTLLVDDTKVTPLTQTKGRNPIETALSHLKGKQFDYSICVLQEDLIX 610

Query: 604 GDKLYLMQIKKCGKGVIVYDKNNGKMSRLWN 634  
 G+KL L+ K +GKV+V + ++ RLK+  
 Sbjct: 611 GKVLNLSHKIRGKVVVNHITNRIIRLGH 641

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 769

A DNA sequence (GBSx0817) was identified in *S. agalactiae* <SEQ ID 2363> which encodes the amino acid sequence <SEQ ID 2364>. This protein is predicted to be RegR (kdgR). Analysis of this protein sequence reveals the following:

```

5 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2545(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:CAB01925 GB:Z79691 RegR [Streptococcus pneumoniae]
 Identities = 222/333 (66%), Positives = 279/333 (83%)

 Query: 1 MSKKQVTINDIAQLSKTSKTTVSFFLNQKFERKMSDETRQRIQEVIDETGYRPTSIARSINS 60
 Sbjct: 1 M KK+TI DIA++++TSKTTVSF+LN K+EKMS ETR++I++VI ET Y+PS +ARSINS
20 KKKLLGLVIGDITNTFSNQIVKGIEHITKQGYQIIIVGNSNYDAKSEEDYIERMNLGLV 120
 Sbjct: 61 K+TKL+GV+IGDITN+FSNQIVKGIE I Q GYQ+++GNSNY +SE+ YIE+ML LGV
 KRTKLGLVIGDITNSFSNQIVKGIEDIASQNGYQVWIGNSNYQSEEDRYIESMLLGLV 120

25 Query: 121 DGFIIQPTSNFRKYSRIKKEKKKPMVFFDSQLYBHKTSWVKANNVDAVIDMTQSCILNGY 180
 Sbjct: 121 DGFIIQPTSNFRKYSRI+ EKKK MVFFDSQLYB+TSWVK NNVDAVIDMTQ C+ +GY
 DGFIIQPTSNFRKYSRIIDEKKKKPMVFFDSQLYBHTSWVKNNVDAVIDMTQSCIEG 180

30 Query: 181 KKFIMITADTSLSTRIERASGFMDALDNGFGYDTLVIEDDDHKSQSDIDFLKAVVPDK 240
 Sbjct: 181 EYFLITADTSLSTRIERASGFVDALTDANRHASLTIEDKHNTLSQIKFLQREIDPD 240
 + F++ITADTS LSTRIERASGF+DAL D + +L IED + I++FL+ +

35 Query: 241 EETLVFAPNCWALPMVFTAMKLNLFDMRVLGVGFDMIEWTDFSSPKVSTIVQPAYEBGE 300
 Sbjct: 241 E+LTVF ENCWALP+VFT +K LN+++P+VGL+GFDM EWT FSSP VST+VQP++EBG+
 EETLVFAPNCWALPLVFTVIKELVNLFPQWGLIGFDMTWTFCSSPSVSTLVQPSFEBG 300

 Query: 301 QVAQILINRIEGBDSDVDNQIVDQCMFWKESTF 333
 Sbjct: 301 Q +LLI++IEG + + QQ++DC + WKESTF
40 QATKILIDIEGRNQERQGVLDSCNNWKESTF 333

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2365> which encodes the amino acid sequence <SEQ ID 2366>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.2928(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

 Identities = 214/333 (64%), Positives = 266/333 (79%), Gaps = 2/333 (0%)

55 Query: 1 MSKKQVTINDIAQLSKTSKTTVSFFLNQKFERKMSDETRQRIQEVIDETGYRPTSIARSINS 60
 Sbjct: 13 M +K+TI DIA+L+KTSKTTVSF+LN +F+KMG+ET+ RI E I T Y+PS ARSUN+
 M+RKVTIDIAELAKTSKTTVSFFLNQKFERKMSDETRQRIQEVIDETGYRPTSIARSINS 72

 Query: 61 KKKLLGLVIGDITNTFSNQIVKGIEHITKQGYQIIIVGNSNYDAKSEEDYIERMNLGLV 120
 Sbjct: 73 K TKL+GV+IGDITN+FSNQIVKGIE ++ GYQII+GNSNYD E+ +R MNLGLV
60 KSTKLIGVIGDITNSFSNQIVGIBSKAQRFQGYQIIIGNSNYDPSKDELEIERMNLGLV 132

```

-872-

Query: 121 DGFIIQPTSNFRKYSRI LKKKKPMVFPDSQLYEHKTSWVKANNMYDAVYDMTQCLNNGY 180  
 DGFIIQPTSNFRKYSRI+ KKK +VFPDSQLYEH+T+VWK NNYDAVYD Q+C++GV  
 Sbjct: 133 DGFIIQPTSNFRKYSRIIDIKKKKVVFPDSQLYEHKTNWVKTNNDYDAVDTIQCCIDRGY 192

5 Query: 161 KKFIMITADTSLSTRIERASQFMALKDNGFGYDTLAVIEDDDHKSIEDFLKGVVFDK 240  
 + FIMIT + LLSTRIERASGF+D L+ N + ++I+++ S I FL+ + K  
 Sbjct: 193 EHFIMITGRNILLSTRIERASGFIDVLEANNHILQRMIDENQTSSEALQFLQSLTKK 252

10 Query: 241 EETLVFAPNCWALPMVPTAMKNNLNTMPRVGLGVFNIEWTDPSSPKVSTIVQPAYEBGE 300  
 +LVF PNCWAL VPTAMK+L P+P +GLGVFNIEWT FSSP ++TI+QPAYEBGE  
 Sbjct: 253 --SLVFPNCWALPKVPTAMKSLKFNIPRIGLVGFNIEWTKFPSSPTITTIQPAYEBGE 310

Query: 301 QVAQILINRIEGBDSDVMQIVDCQMPKKESTF 333  
 Q +ILI+ IEG QQI DCQ+ N+ESTF  
 15 Sbjct: 311 QATKILIDITEGHSQEAQQIIFDCQVNWQESTF 343

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 770

20 A DNA sequence (GBSx0818) was identified in *S. agalactiae* <SEQ ID 2367> which encodes the amino acid sequence <SEQ ID 2368>. This protein is predicted to be polypeptide deformylase (def-1). Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2339 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC15392 GB:AJ278785 polypeptide deformylase (Streptococcus pneumoniae)  
 Identities = 169/204 (82%), Positives = 192/204 (93%), Gaps = 1/204 (0%)

35 Query: 1 MSALDKIAKASHLIDNDIIREGNPTLRKVAEEVTFPLSEKEILGERMNOFLKHSQDPI 60  
 MSAI+++ KA+HLIDNDIIREGNPTLR+AKEVTFPLS++R ILGERMNOFLKHSQD+  
 Sbjct: 1 MSATERTTGAHLIDNDIIREGNPTLRATAEVTFPLSDQRIILGERMNOFLKHSQD+ 60

40 Query: 61 MAEKGLRGVGLAAPQLDISKRIIAVLVFNVEDAQGNPPKRAYSLQEVMTNPKVSHSV 120  
 MAEK+GLRGVGLAAPQLDISKRIIAVLVFN+ +G P+RAY L+ +MYNPK+VSHSV  
 Sbjct: 61 MAZKGLRGVGLAAPQLDISKRIIAVLVFNIVE-EORTQRAYLEALMYNPKVSHSV 119

45 Query: 121 QDAALSDGEGCLSDREVPGYVVRHARVTIEYFDKTEGKHRLKLGYSINVQHEIDHID 180  
 QDNL +GEGCLSDVR VPGYVVRHARVT++YFDK GEGH+KLGYSINVQHEIDH+  
 Sbjct: 120 QDAALGEGCLSDVRNPGYVVRHARVTVDYFDKDGEGKHRIKLGYSINVQHEIDHIN 179

Query: 181 GIMFYDRINEKNPFAVKEGLLILE 204  
 GIMFYDRINEK+PFAVK+GLLILE  
 Sbjct: 180 GIMFYDRINEKDPFAVKDGLLILE 203

50

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2369> which encodes the amino acid sequence <SEQ ID 2370>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1745 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 160/204 (78%), Positives = 186/204 (90%)

```

5 Query: 1 MSAIDKLKVKASHLIDMNDIIREGNPTLRKVAEEVTFPLSEKREILGKMMQFLKHSQDPI 60
 MSA DKL+K SHLI M+DIIREGNPTLR VA+EV+ FL +++ +LGEKMMQFLKHSQDP+
 Sbjct: 1 MSAQDKLKPKSHLITMDDIIREGNPTLRKVAKEVSEFLCEDILLGKMMQFLKHSQDVF 60

10 Query: 61 MAEKLGIRGVGLAAPQLDISKRIITAVLVPHVEDAQGNPPKRAYSLQEVYNNPKIVSHSV 120
 MAEKLGIR GVGLAAPQ+D+SKRIITAVLVPH+ D +GNPPKRAYS QEV+YNNK+VSHSV
 Sbjct: 61 MAEKLGIRAGVGLAAPQIDVSKRIITAVLVPHLPDKGKNPPKRAYSWQEVLYNNPKIVSHSV 120

15 Query: 121 QDAALSDGEGCLSDVREYVPGYVVRHARVTIETFDKTGEKHLKLGKYN+IVVQHEIDHID 180
 QDAALSDGEGCLSVDR V GYVVRHARVT++Y+DK G+HR+KLGKYN+IVVQHEIDHI+
 Sbjct: 121 QDAALSDGEGCLSDVRVGVGVVRHARVTVDYDREGQQHRIKLGKYN+IVVQHEIDHIN 180

 Query: 181 GIMFYDRINEKNPFAVKEGLLILE 204
 G++FYDRIN KNPF KE LIL+
 Sbjct: 181 GVLFYDRINAKNPFETKEELLILD 204

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 771

A DNA sequence (GBSx0819) was identified in *S. agalactiae* <SEQ ID 2371> which encodes the amino acid sequence <SEQ ID 2372>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
30 bacterial cytoplasm --- Certainty=0.3620(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10177> which encodes amino acid sequence <SEQ ID 10178> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC75224 GB:AR000305 putative transcriptional regulator
[Bacterichia coli K12]
Identities = 58/191 (30%), Positives = 98/191 (50%)

40 Query: 37 DLQVITLTAGQSVCKQSGQLVLYHYIVKGRPKIVRRLFNGKSHILDIKTKPYLIGDIELL 96
 D++ A + ++G+Q +L Y+ +GR ++ L NG+ ++D P IG+IEL+
 Sbjct: 17 DTRLPHFLARDYIVQSGQPSWLFYTLRGRARLYATLANGRVSLDFFPAAPCFIGSHIELL 76

45 Query: 97 TNRQIVSVIALEDLTVIQLSLKGRKEKLLTDATFLKLKSGELAQPHDQNKASTNLGY 156
 +V A+E+ + L +K + LL D FL KL L+ + + N +
 Sbjct: 77 DKDHEPRAVQATEEONCALFMKHYRPLLLNDTLFLRKLCTLSHKNYNTVSLTQNGSF 136

50 Query: 157 TVKELLASHILATEEQYFQLELSLSDSPGVSVYRHLLRVIHDMVKEGLIKRQKPYFPIK 216
 + LA+ LL +E + + + A+ GVSYRHLL V+ + +GL+ K K Y IK
 Sbjct: 137 FLVNRILAAPILLSQESDLYHRKHTQAARLYGSVYRHLLVLAQFTHIGLLIKSKGYLIK 196

 Query: 217 NRPFALESINIQ 227
 NR L L ++
55 Sbjct: 197 NRRQSLGLALE 207

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2373> which encodes the amino acid sequence <SEQ ID 2374>. Analysis of this protein sequence reveals the following:

-874-

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3809(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 23/63 (36%), Positives = 35/63 (55%), Gaps = 1/63 (1%)  
 Query: 146 QNFKASINLGYTVKELASHILAEIEGYPQLKSLSLADSPGVSYRHILARVHDMVKSGI 205  
 GN+ N+ YTVKE AS+ L + L L+ LA+ PG S RHL V+ + + +  
 15 Sbjct: 3 QNV-CQQNITTYTVKERFASYLEGAQANQEVHLNITLLANRFCTSDRHLHVLKQIPQRI 61  
 Query: 206 IQK 208  
 I++  
 Sbjct: 62 IER 64

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 772

A DNA sequence (GBSx0820) was identified in *S. agalactiae* <SEQ ID 2375> which encodes the amino acid sequence <SEQ ID 2376>. Analysis of this protein sequence reveals the following:

25 Possible site: 54  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -9.24 Transmembrane 163 - 179 ( 159 - 185)  
 INTEGRAL Likelihood = -8.49 Transmembrane 204 - 220 ( 201 - 226)  
 INTEGRAL Likelihood = -7.80 Transmembrane 272 - 288 ( 269 - 296)  
 30 INTEGRAL Likelihood = -6.00 Transmembrane 333 - 349 ( 331 - 352)  
 INTEGRAL Likelihood = -5.41 Transmembrane 75 - 91 ( 73 - 92)  
 INTEGRAL Likelihood = -4.94 Transmembrane 245 - 261 ( 240 - 262)  
 INTEGRAL Likelihood = -4.41 Transmembrane 362 - 378 ( 359 - 380)  
 INTEGRAL Likelihood = -4.14 Transmembrane 96 - 112 ( 95 - 113)  
 35 INTEGRAL Likelihood = -2.44 Transmembrane 141 - 157 ( 141 - 158)  
 INTEGRAL Likelihood = -1.81 Transmembrane 302 - 318 ( 301 - 320)  
 ----- Final Results -----  
 40 bacterial membrane --- Certainty=0.4694(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8659> which encodes amino acid sequence <SEQ ID 8660> was also identified. Analysis of this protein sequence reveals the following:

45 Lipop: Possible site: -1 Crend: 7  
 MG: Discrim Score: -3.52  
 GVH: Signal Score (-7.5): 0.340001  
 Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 50 ALOM program count: 11 value: -9.24 threshold: 0.0  
 INTEGRAL Likelihood = -9.24 Transmembrane 134 - 150 ( 130 - 156)  
 INTEGRAL Likelihood = -8.60 Transmembrane 17 - 33 ( 13 - 37)  
 INTEGRAL Likelihood = -8.49 Transmembrane 175 - 191 ( 172 - 197)  
 INTEGRAL Likelihood = -7.80 Transmembrane 243 - 259 ( 240 - 267)  
 55 INTEGRAL Likelihood = -6.00 Transmembrane 304 - 320 ( 302 - 323)  
 INTEGRAL Likelihood = -5.41 Transmembrane 46 - 62 ( 44 - 63)  
 INTEGRAL Likelihood = -4.94 Transmembrane 216 - 232 ( 211 - 233)  
 INTEGRAL Likelihood = -4.41 Transmembrane 333 - 349 ( 330 - 351)  
 INTEGRAL Likelihood = -4.14 Transmembrane 67 - 83 ( 66 - 84)  
 60 INTEGRAL Likelihood = -2.44 Transmembrane 112 - 128 ( 112 - 129)



-875-

INTEGRAL Likelihood = -1.81 Transmembrane 273 - 289 ( 272 - 291)  
 PERIPHERAL Likelihood = 3.45 193  
 modified ALOM score: 2.35

5 \*\*\* Reasoning Step: 3

----- Final Results -----

10 bacterial membrane --- Certainty=0.4694(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA550057 GB:AJ240286 TRANSPORT PROTEIN, permease [Pyrococcus abyssi]  
 Identities = 94/382 (24%), Positives = 173/382 (44%), Gaps = 30/382 (7%)

15 Query: 5 NEKLSSLSL-SLILLSTFTSPALPQMISY-RDKGLPSQVLELLSPISPMALIFILLIT 62  
 MEKL +L L SL + +S A+P + +D G+ + + LL + + I +  
 Sbjct: 1 NEKLILILISIGWIFNYSHRMVPSLAPIIMKDLGINNAEIGLLMTSLLPYSLIQVPA 60

20 Query: 63 PWLSKKLSEKHIIIFGLLLTALGGGLPVVSNVLLVPSRLLGSGIGFINTRAVISIE 122  
 ++ K+ K ++ +L +L L V+++Y + R L G G A ++ISE  
 Sbjct: 61 GTYIDKIRKKLLTISILGYLSLALIVLTRDYDWLVTRALYGFAGLYYAPATALISE 120

25 Query: 123 YTGKERRKLLGLRGSFEVLGNA---GLTAL--VGLLLTFQWSKSPMIVFLALPIVLNL 177  
 ++ ++ L F ++G A G+T L V + LT W +P++ + I+ L  
 Sbjct: 121 LFRKRKGSAL-----GPFMVGAIGSGITPLIVVPVALTSMRYAFVLVLSIMSIIVGILL 175

30 Query: 178 VFAPKVVKTNDIKTKGOKIPKADLTIVALAILAGFVITITGINLRIPILVVEPLG 237  
 + A K + IK +G K +++++LA G + + LW G+  
 Sbjct: 176 MVAIK-----GEPIKVEGVKFIKIPRGVFLLSANFLISLGAFFAM-LTFVLVSVLR-GV 227

35 Query: 238 GTPAQSLSLSSAMMLGIAGNPSGOLAMFHKQIPICLVLFSS-LTLGLVGLSPSLMVL 296  
 G +SL+ S + L+GI+ + G L K + + L S LT L + +PS L ++  
 Sbjct: 228 GME-KASLMFSMLSGLVGLSGIAGFLYDHLGKSVLLAYALMSLLTFLVIVISPLPLI 286

40 Query: 297 TISMASGFLYSI--MVTAVFSLVADRVEYSLVGSATTLLVLF--CHIGGASAILSSCFD 353  
 + + LYS+ ++TA S A R +V +V F IG L+  
 Sbjct: 287 PLGLV----LYSVGGIMTAYTSEKASRENLGVMGFVMMVGFPGATIGPYIVGFLIDRLG 342

40 Query: 354 HLLQQINAVFFVYAILSLAVGM 375  
 + L + +V Y + + +G+  
 Sbjct: 343 YSLALL-SVPLATLVSAVIIGL 363

There is also homology to SEQ ID 2378.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 773

A DNA sequence (GBSx0821) was identified in *S. galactiae* <SEQ ID 2379> which encodes the amino acid sequence <SEQ ID 2380>. Analysis of this protein sequence reveals the following:

50 Possible site: 23  
 >> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.38 Transmembrane 171 - 187 ( 171 - 187)

55 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-876-

>GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces coelicolor A3(2)]  
Identities = 101/327 (30%), Positives = 169/327 (50%), Gaps = 12/327 (3%)

5 Query: 8 WATLGTVIANEL-AQALEARQCKLYSVANRTYDKGLBFAATKYGIQKYYDHDQVPEPE 66  
W L TG +A A ++ ++ +VA+RT FA ++GI + Y + + D +  
Sbjct: 11 WGILATGGAARFTADLDLPDAEVVAVASRTASAKTFAERFGIPRAYGGWETLARDED 70

10 Query: 67 VDIYISTPHNTHISFLKALANGKIVLCEKSTITNSELKEAIDLAEITHVLAEMTI 126  
VD++Y++TPH+ H + L G+VLCCK TIAN+ E E + LA N V L EM +  
Sbjct: 71 VDVVYVATPHSAHRTAAGLCLEAGRNVLCEKPTFLNAREAAELVALARENGVFLMENEMW 130

15 Query: 127 PHMPTYRQLKTLVDSCKLGLPMQMNFGSYKIKYDWINRFFSRDLAGGALLDIGVYALSC 186  
+ P+ R+LK LV G +G ++ +Q +FG + +R GGALLD+GVY +S  
Sbjct: 131 YCNPLVRRLKELVADGAIGSEVRSLQADFGLAGFPFAHRLRDPACGGALLDLGVYVVSF 190

20 Query: 187 IRWFMSSEAPHNITSQVTFAPITGVDEQVGLLTNDPANEMATVSLHAKQPKRATIAIDK 246  
+ + E P ++ ++ + GVD Q G LL+ + +A++ S+ P A+I +G  
Sbjct: 191 AQULLGE-PTDVAARVLSSEBGLDTQGLLSYGDALASIHCSITGTSNBSAITSSEB 249

Query: 247 YIEU---FEYPRGQKAVITYTREGHDDIL--EAGKTENALQYEVADMERAV-SGKINH-- 298  
I++ F+P V+ T Q+ A +L++E ++ A++G+T  
Sbjct: 250 RIDVNGSFFFP--DHFVLRITGRDPQEFRADPADGPRESLHSEAEVMRALRAGHTESPL 307

25 Query: 299 MYLNTYKVDMDIMTQLRQWGFYTEE 325  
+ L+ T VM + +R G YP E  
Sbjct: 308 VPLDGLTLMVRLTDAIRDRVGKYPGE 334

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
30 vaccines or diagnostics.

### Example 774

A DNA sequence (GBSx0822) was identified in *S. agalactiae* <SEQ ID 2381> which encodes the amino acid sequence <SEQ ID 2382>. This protein is predicted to be oligopeptidase. Analysis of this protein sequence reveals the following:

35 Possible site: 19  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
40 bacterial cytoplasm --- Certainty=0.2681(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC14579 GB:A249396 oligopeptidase [Streptococcus thermophilus]  
Identities = 504/631 (79%), Positives = 563/631 (88%)

45 Query: 1 MIKYQDDFYQAVNGEAKTAVIPDDKPRITGGPSDLADDIEALMLSTTDKHLANDKPSDT 60  
M + QDDFY A+NGEW KTAIVIPDDK TGGPSDIAD+IE LML TTD+MLA EN P +  
Sbjct: 1 MTRLQDDFYHAINGEAKTAVIPDDKPRITGGPSDLADEIDMLITTDQMLAGENVPINA 60

50 Query: 61 ILNHFIAPRMTADYQKREEVGVSPVPLIRYKGLQSPSEFASKVAEYELSGKPNRFF 120  
IL +PI FH+MTADY +RE VG+ PV PLIRYK L SPSEFASK+AEY+ GKPNRFF  
Sbjct: 61 ILQNFIKPHRMTADYDRREAVGIEPVKPLIRYKGLSSPSSEFASKIAGYMSGKPNRFF 120

55 Query: 121 GVAPDFMNAQLNVAEADPGIILPDTTYTSHNEKGKRLIAFWRESQDILLPLGLSEGE 180  
V+PDFMNAQLNVAEADPGIILPDTTYTSHNEKGKRL WR+ QE+LL +G + +E  
Sbjct: 121 SVSPDFMNAQLNVAEADPGIILPDTTYTSHNEKGKRLIAFWRESQDILLPLGLSEGE 180

60 Query: 181 IKDILDKVLDLAKLAQYVLSRESSEYVKLYHPYNWEDFTKLAPELPLDAI FQKILGQK 240  
IKD+LDK+ LQKLA+YVLS RESSEY+LYHPY+W DFTKLAPELPLD+IF +ILGQ  
Sbjct: 181 IKDILDKVLDLAKLAQYVLSRESSEYVELYHPYDADFTKLAPELPLDSIFTEILGQV 240

-877-

Query: 241 PDKVIVPEERFWTEFASDYSESNMELLKADLILSAANAYNAYLTDIDIRIKSGVYSRALS 300  
 PDKVIV ES FWTEFA++YYSE+NMELLKA L++ A ++NAYLTD++R+ SG YSRALS  
 Sbjct: 241 PDKVIVSESPFWTEFAAAYSESNMELLKAVLLIDATTSNAYLTDIELRVLSGGYSRALS 300

5 Query: 301 GTPQAMDKKKAAYTLASGPNQALGLWYAGEKFSPEAKADVEHKIATMIDVYSKRLKAD 360  
 GTPQAMDKKKAA+YLA GPNQALGLWYAGEKFSPEAKADVE K+ATMIDVYSKRI+ AD  
 Sbjct: 301 GTPQAMDKKKAAYTLAQSPYNQALGLWYAGEKFSPEAKADVEAKVATMIDVYSKRLQAD 360

10 Query: 361 WLAQSTREKAIKINAVITPHIGYPEKLPETVTKKIDPKLSLVENATMLKIKIAYKWSK 420  
 WLA TREKAI KINAVITPHIGYPEKLPETV TKKID LSLVENA L +ISIA+ WSK  
 Sbjct: 361 WLAPETREKAITKINAVITPHIGYPEKLPETVTKKIDENLSLVENAKVLVETSIASWSK 420

15 Query: 421 WNKPVDRSEHMPAHMVNAYDPOQNIQVFPAILQEPFYALBOSSSANYGGIGAVIAHE 480  
 WNKPVDRSEHMPAHMVNAYDPOQNIQVFPAILQ PFY + QSSSANYGGIGAVIAHE  
 Sbjct: 421 WNKPVDRSEHMPAHMVNAYDPOQNIQVFPAILQAPFYDLAQSSSANYGGIGAVIAHE 480

20 Query: 481 ISHAFDTNGASFDEHGLSNWMTDEDPEAPKKLTKRVQFDGLESYGAKVNGKLTUSEN 540  
 ISHAFDTNGASFDE+GSL NWMT++D+ APK+ TDK+V+QF+GL+SYGAKVNGKLTUSEN  
 Sbjct: 481 ISHAFDTNGASFDENGSLQWMTEDDYAAPKERTDKIVQFBSGLDSYGAKVNGKLTUSEN 540

25 Query: 541 VADLGGVACALEAAQRESDFSRDFFINPATINRMKARDEYMCMLASVDVHAPAKWRIN 600  
 VADLGGVACALEAA++R+ DFS R+FFINPATIWR KAR+EYMCMLASVDVHAPAKWRIN+  
 Sbjct: 541 VADLGGVACALEAAKRESDFSVREFFINPATIWRKAREYMCMLASVDVHAPAKWRIN 600

Query: 601 TVTNFEFPHKEFDVKGDNMRPVEKRVII 631  
 TVTN+EFPHKEFDVK+GD MMR E RVIIW  
 Sbjct: 601 IVTNFDEPHKEFDVKEGDGMWRAPEDRVII 631

30 Endopeptidases are often exposed antigens.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2384> which encodes the amino acid sequence <SEQ ID 2384>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2622 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 504/631 (79%), Positives = 564/631 (88%)

Query: 1 NIKYQDDFYQAVNGBNKATVIPPDKPRTGGPSLADDIETALMLSTTDKVLADENKPSDT 60  
 N YQDDFYQAVNG+NA+TAVIPDDKPRGGPSLAD+ETALML TTD WLA EN P D  
 Sbjct: 1 NHTYQDDFYQAVNGKNATVAVIPDDKPRGGPSLAD+ETALMLDITDAMLAGENIPDA 60

Query: 61 ILNHPIAPHIRKADYQREKAVNVSFVPLFLIEEYKLGQSPSEFASKVAVSEYLGKPIEFPP 120  
 IL +F+ FH++ ADY KR+EVGVSP+LPLIEEY+ L+SPSEF +A+YEL G ERIEFP  
 Sbjct: 61 ILKNFVKFHRLLVADYAKRDEGVGSPLFLIEEYQSLKSPSEFVANTAKYKDLGLPIEFPP 120

Query: 121 GVAPDFPNAQLNVLNWAEPGILLPDTTYSSENEKQKELLA+WRKQSQDILLPLGLSBSQS 180  
 VAPDFPNAQLNVLNWAEP L+LDPDTTYY E NKK +EL WR+SQS LLP PG S +E  
 Sbjct: 121 SVAPDFPNAQLNVLNWAEPGILLPDTTYSSENEKAEELRGIWRQSQKLLPQPGSTEK 180

55 Query: 181 IKDILDKVILDAKLAQVLSRESESEYVKLYHFMWEDFTKLAPLPLDAIPQKILQOK 240  
 IKD+LDKVA LD +LA+VYLSREE SEY KLYHFP W DF KLAPLPLD+IP+KILQ  
 Sbjct: 181 IKDILDKVIELDKLAQVLSRESESEYAKLYHFPVWADFKKLAPLPLDSIFEKILQOV 240

60 Query: 241 PDKVIVPEERFWTEFASDYSESNMELLKADLILSAANAYNAYLTDIDIRIKSGVYSRALS 300  
 PDKVIVPEERFWTEFA+ YYSE+NM+LLKA+LI+ AANAYNAYLTDIDIR+SG YSRALS  
 Sbjct: 241 PDKVIVPEERFWTEFATYSESNMELLKANLIVDAANAYNAYLTDIVESGYSRALS 300

Query: 301 GTPQAMDKKKAAYTLASGPNQALGLWYAGEKFSPEAKADVEHKIATMIDVYSKRLKAD 360

-878-

GTPQAMDK+KAA+YLA GP++QALGLWYAG+KPSPEAKADVE K+A MI+VYKSRLE AD  
 Sbjct: 301 GTPQAMDKQAAPFLAQGPFSQALGLWYAGQKPSPEAKADVESKVARMEIVYKSRLETDAD 360  
 Query: 361 WLAQSTREKAIKMNIVITPHIGYPEKLPETYTKKIIDPKLSLVENATNLKISIAVGSK 420  
 5 WLA+TRKAI KMNIVITPHIGYPEKLPETY KK+ID LSLVENA NL KI+IA+ WSK  
 Sbjct: 361 WLA PATREKAITKLNIVITPHIGYPEKLPETYAKKVIDESLSLVENATNLKITAHTWSK 420  
 Query: 421 WNKPVDRSENMIPAHVWYAYDQCQNIQVFPAAILQEPYALEQSSSANYGGIGAVIAHE 480  
 10 WNKPVDRSENMIPAHVWYAYD QCQNIQVFPAAILQEPY+L+QSSSANYGGIGAVIAHE  
 Sbjct: 421 WNKPVDRSENMIPAHVWYAYDQCQNIQVFPAAILQEPYFSLDQSSSANYGGIGAVIAHE 480  
 Query: 481 ISHAFDTNGASFDEHGSILNNWITDEFEAFKLTIDVVBQFDGLSEYGAKNVGLTVSEN 540  
 ISHAFDTNGASFDEHGSILN+WIT ED+ AFK+ TDK+V QFDGLSEYGAKNVGLTVSEN  
 15 Sbjct: 481 ISHAFDTNGASFDEHGSILNNWITQEDVAAFKERTDKIVAQFDGLSEHGAKNVGLTVSEN 540  
 Query: 541 VADLGGVACALEAAQESDPSARDFFINFATINRMKARDEYMQMLASVDVHAPAKMTNI 600  
 VADLGGVACALEAAQ E DPSARDFFINFATINRMKAR+EYMQMLAS+DVHAP +RTN+  
 Sbjct: 541 VADLGGVACALEAAQESDPSARDFFINFATINRMKAREYMQMLASIVDHAPGLKNTV 600  
 20 Query: 601 TVTNFSEPHKEFDVDEGDMWRAPVEKRVIIW 631  
 T+TNF+ PH+ FD+K+GD MWR +RVIIW  
 Sbjct: 601 TLTNFDPHETFDIKEGDMWRAPKDRVIIW 631

SEQ ID 2382 (GBS193) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell  
 25 extract is shown in Figure 23 (lane 3; MW 73kDa).

The GBS193-His fusion product was purified (Figure 196, lane 5) and used to immunise mice. The  
 resulting antiserum was tested for Western blot (Figure 253). These tests confirm that the protein is  
 immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 30 vaccines or diagnostics.

### Example 775

A DNA sequence (GBSx0823) was identified in *S. agalactiae* <SEQ ID 2385> which encodes the amino  
 acid sequence <SEQ ID 2386>. This protein is predicted to be immunity protein (mccF-1). Analysis of this  
 protein sequence reveals the following:

35 Possible site: 36  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 40 bacterial cytoplasm --- Certainty=0.1627 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9433> which encodes amino acid sequence <SEQ ID 9434>  
 was also identified.

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AA884435 GB:AF027868 YocD [Bacillus subtilis]  
 Identities = 114/270 (42%), Positives = 170/270 (62%), Gaps = 4/270 (1%)  
 50 Query: 1 MSPSGHYLNDILYSASITSRVDELHEAPADPSVDIALIATIGGPNENKLLPYLDYDLISK 60  
 ++ +H E + S+SI SRV DLH AP DP V AIL T+GGPNEN+LL YLDY+ I +  
 Sbjct: 43 VTIAEHAENECNEFDSSSISERVDLHAAFPDGVKAILTTLGGPNENQLRLYDYEKIKR 102  
 Query: 61 NPKIICGYSDSTAFNALNPAKAKIQTYMGDAYSSFKMKKEGQPYCTQAWLT-AMTEGHYEL 119  
 +PKI+CGYSD TA NAI+ K + TY GP +S+F MK+G Y + +L+ ++ +E+  
 55 Sbjct: 103 HPKILCGYSDITALCNALYQKTLGLVTSYSGHPSTFAMKKGIDYTEEYPLSCASDDPPEI 162

Query: 120 WPSEWSSDPWYDPSKPRQFFPTWK-IYNEHGASGTIIGNLTGGLRGTPYAPKIER 178  
 PS EMS D W+ + R+F+P + G A GT+IGNLT LL+GT Y P+ E  
 Sbjct: 163 HPSEWSSDRLWLDQENRRFYPNKGVVQVQSYAEGTLIGNLTCTNLGQTEYFPTEH 222

Query: 179 YVLLIEAEESNIFYEDFNLAII--LQAYPHQAILMGRFPKCGMTPQVPSYILSKHAI 236  
 +LLIE+ S+ + FDR+L++ L A+ H +ALL+GRF K ++ + + ++  
 Sbjct: 223 TILLIEEDYSDIHMFDRLLQSLHLPAFSAVKILIGRFQKASNVSDLVKMIETKKE 282

Query: 237 FKEIPVIYDMDFAHQTPLLTVTIGAEISVD 266  
 IP+I +++ HT P+ T IG ++  
 Sbjct: 263 LSGPIILANINAGHTSPITATFPIGGTCRIE 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2387> which encodes the amino acid sequence <SEQ ID 2388>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1162 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 75/252 (29%), Positives = 125/252 (48%), Gaps = 22/252 (8%)

Query: 34 VDAILATIGGFNSNELLFYLDVLLISKNPKICGYSDSTAPLNAI FAKAKIQTVMGPAYS 93  
 VD I+ +IGG+NSN +L Y+DYDL + I GYSD+TA A++ K TY+ +  
 Sbjct: 1 VDVINTSIGGYNSNMLYKIDYDLFKQKFPFPIGYSDTTALALALYKTCITVLSQSVI 60

Query: 94 SFYMKGEQGP-----YQTAQMLTAMTENHYELNPSEWSSDPWYDPSKPRQFFPTE 143  
 S E+P + Q+ + ++W++EW + W ++ ++ E  
 Sbjct: 61 S-NFGEFEPFNLNIFYFDMLQSKCETLMVQIPDVW--TDWIN--METYERTKKINKNE 116

Query: 144 WKIYNHGKASGTIIGNLTGGLRGTPYAPKIERIYVLLIEAEESNIFYEDFNLA--AI 201  
 W I+N G+ +GT+IGNLT + GT Y PKI +L+ E ++ RN A+  
 Sbjct: 117 WIIPNKGEPNLTIGCNLDYVIGIOTEXMPKITEDTILLIEDVYTDLGRLYRNFTTAL 176

Query: 202 LQAYPHQAILMGRFPKCGMTPQVPSYILSKHAI FKEIPVIYDMDFAHQTPLLTVTIGA 261  
 + +++ +P + G V I++ ++IP++ + D HT P + IG  
 Sbjct: 177 HGIFDKIGGLIISK-P-ETIGENSVDVINDIINEFVGHKIPILINFDGHTHPSCLAPIGG 235

Query: 262 ELSVDVTTLSLS 273  
 ++ TLSLS  
 Sbjct: 236 KI-----TLSLS 242

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 776

A DNA sequence (GBSx0824) was identified in *S.agalactiae* <SEQ ID 2389> which encodes the amino acid sequence <SEQ ID 2390>. Analysis of this protein sequence reveals the following:

possible site: 15  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3112 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 777

- 5 A DNA sequence (GBSx0825) was identified in *S.agalactiae* <SEQ ID 2391> which encodes the amino acid sequence <SEQ ID 2392>. Analysis of this protein sequence reveals the following:

Possible site: 21  
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.6171 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 10175> which encodes amino acid sequence <SEQ ID 10176> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

- 20

#### Example 778

A DNA sequence (GBSx0826) was identified in *S.agalactiae* <SEQ ID 2393> which encodes the amino acid sequence <SEQ ID 2394>. Analysis of this protein sequence reveals the following:

Possible site: 19  
>>> Seems to have an uncleavable N-term signal seq  
INTEGRAL Likelihood = -10.19 Transmembrane 83 - 99 ( 80 - 113)  
INTEGRAL Likelihood = -9.71 Transmembrane 4 - 20 ( 1 - 24)  
INTEGRAL Likelihood = -9.45 Transmembrane 315 - 331 ( 307 - 337)  
INTEGRAL Likelihood = -8.33 Transmembrane 186 - 202 ( 180 - 210)  
INTEGRAL Likelihood = -7.75 Transmembrane 233 - 249 ( 227 - 255)  
INTEGRAL Likelihood = -3.98 Transmembrane 390 - 406 ( 382 - 407)  
INTEGRAL Likelihood = -3.61 Transmembrane 27 - 43 ( 27 - 45)  
INTEGRAL Likelihood = -3.29 Transmembrane 107 - 123 ( 105 - 125)  
INTEGRAL Likelihood = -1.75 Transmembrane 273 - 289 ( 273 - 290)

- 35 ----- Final Results -----  
bacterial membrane --- Certainty=0.5076 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15347 GB:299121 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 174/524 (33%), Positives = 275/524 (52%), Gaps = 13/524 (2%)

- 45 Query: 1 MESTILVSLFLILSLVINRIFPKPLPFIQLVFGILSGLVFKHSQVHIDPELFLAF/ 60  
M+ +++ L + +SN++NR P +P+P IQ+ GIL+ ++ ELF  
Sbjct: 1 MDPIFVLVLLITIAISINVINRIFPIFVPLIQVALGILAAFPQGLHFLNTELFFVLV 60
- Query: 61 IAPIAFREGQESDIFSFKYRAILYLILPTVLTALVGVGVGHLLPVSPLAACFALG 120  
IAPL P +G+ + RA IL L L VF T IV GY ++P ++PLAA F L  
50 Sbjct: 61 IAPLLPNDGKRTPRALNLRAPILLALGLVFATVIVGGYTIHWMP-AIPLAAAFGLA 119

-881-

Query: 121 AALGPTDAVAFISIAKRFQFPKRAKNILKLEGLINDASGLVSPQFALTAIVGYFSLAKA 180  
 A L PTD VA +++ R + PK +L+ EGL+NDASGLV+P+FA+ A VTG FSLA+A  
 Sbjct: 120 AILSPDQVAVSALSRVMPKGIILRLLEGLMNDASGLVAPKFAIAAAVTGAFSLAQ 179

5 Query: 181 SLKLALAIMCGFLIGLLFALMRLCLTVLEKFDADVTGALLLELILPFAVYFADLLGF 240  
 ++ +GG L G++ +FL+ L + DVT +L+++ PFV Y A+ +G  
 Sbjct: 180 AVSFVFISLGLGVSIVSFLIIRFRLFLRLGMDQVTFMMLIQILTPFVITLAASRIGV 239

10 Query: 241 SAIIVVAVGVQANRLKCVTLFDAQVDRTSVINFTLNFILNGLVFLIPGRELTRIGP 300  
 S I+AVV G+ A ++ ++ V+S W + FILNGLVF+I G ++ +I  
 Sbjct: 240 SGTLAVAGGITHAVEQDRLESTNKLQIVSSSTWNILFILNGLVFLIPGVTQIPWISV 299

15 Query: 301 LLZSNAYSNFDLISIVLAVTCTFLVRFVAVSCFY--AWRSFKYHKSFKKYWRRIQLAT 358  
 + A SN +I +++T TL L+RFL V F+ W K +K R L++  
 Sbjct: 300 IPRDTAISNMKVIQYILVITFTMLLRPLWLVFPWNGKFFPKDQNIYKQLRSTLLISI 359

20 Query: 359 SGVKSVSIIATILLPKHSVIGE--IGYSLLIFVGAVTMSPLTGLLVLPKLAFLQVK 416  
 SGV+G+V++A +P G +LILF V L + + +VLE L +  
 Sbjct: 360 SGVRGAVTLAGSPSIPYPLEDGTFFPERNLILFLAAGVILCTLVIAIVVLEPILTEKREED 419

Query: 417 DD-----YLIRLSILTKVLSVLEDEKGSSENQASFYAVIDVNSIRHLILEQ--ESSDI 469  
 ++ R ++ L +ED + AS AVI YN +++L +O S I  
 Sbjct: 420 KERNKLLTARKLKIATLQTIKEDMNETNKIASL-AVIAEYNERKMNLRPQQTSSNRI 478

25 Query: 470 KDLDELQNMNLSIESDGLAAYRYGNISIKERYIYQRYLKYLE 513  
 KK +++ + E + L G+I + + O LE  
 Sbjct: 479 KGERKRVAGVGKAEQSAIMKMLRGDIPSTANVLQERFNELE 522

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 779

A DNA sequence (GBSx0827) was identified in *S.agalactiae* <SEQ ID 2395> which encodes the amino acid sequence <SEQ ID 2396>. Analysis of this protein sequence reveals the following:

35 Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

---- Final Results ----  
 bacterial cytoplasm --- Certainty=0.3494 (Affirmative) < succ>  
 40 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 780

A DNA sequence (GBSx0828) was identified in *S.agalactiae* <SEQ ID 2397> which encodes the amino acid sequence <SEQ ID 2398>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

---- Final Results ----

-882-

bacterial cytoplasm --- Certainty=0.5094 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 10173> which encodes amino acid sequence <SEQ ID 10174> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF12706 GB:AF066865 integrase [bacteriophage TPW22]  
 Identities = 171/353 (48%), Positives = 253/353 (71%), Gaps = 1/353 (0%)

10 Query: 21 MASVYKRENGLWEYRISVKTIDQKVRKKGQFQTKKLAQAARIRKKTQNLINDEIV 80  
 MA++RKR W++R+SYK +G+YK+ EKSG+RTKK A+AAA E+K+L + +++++  
 Sbjct: 1 MANFAKRGK-TWQFRLSYKDNNGYKFKKGGYKTKKEEAANDEAKRLNHHISEPNDI 59

15 Query: 81 TLNDFVKTWSEVYKRPYVVDKWTWYSGNPKHKNYFQELKVKDITPLYQKKLNEFGEK 140  
 +LYDF + W++VYK+P+V + TW TY + I Y ++ + +ITP +YQ LN+  
 Sbjct: 60 SLYDFFEKWAKVYKKPHVTKATWRTYKRTLNLDIKYIKDKPIAETPTPTFYQAVLNKMSLL 119

20 Query: 141 YAQETLEKPHYQIKGMKVAVREQVYTPNFAGAKVKVQVEPKNEEDPLEEREYKALLA 200  
 YQE+LAKP++QIK AMK+AV E+H+ NFA+ K KS++ + EE +L EY LLA  
 Sbjct: 120 YRQESLCKFPYQIKSAMKIAVHEKVISENFADPTAKSKLAARFVEEYKLAHDEYKLLA 179

25 Query: 201 LTRNNIQVYSYPTLYLLAVTGLRPSBAWGLTWSDIDFKGILDINKSPDYNTQDFADLK 260  
 + E ++Y SYF YL AVTG+RF+E +GHWG +DF + I +++DYS T +FA+ K  
 Sbjct: 180 LAEEKMEYTSYFACYLTAVTGMRPAELLGLINSHVDFDKKEISQRTWDYSITNNFAETK 239

30 Query: 261 NESSKRKRVIDSNTIDILREYKKGHWQANTIKRVCVGVSNACNKLKKIVGRKVRNHS 320  
 NESSKRK+PI S TI +L++YKK +W N +RV + +SN+ NK IK I GRKV HSL  
 Sbjct: 240 NESSKRKPISSKTIKLLKYYKKEYWHENKYDRVYTNLNNGLNKTIKVLNKRKVHPSL 299

Query: 321 RHTYASFLINGVDIVTISKLLGHESPDITLVYTHQMEALERNPKIKNTIF 373  
 RH++AS+LI G+D+T+S KLLGHE+ ++TLKVY HQ++ + + N + I+ IF  
 Sbjct: 300 RHFASVYLIYKIDLLTVSKLLGHENLAVTLKVYHQLKEMGEQNDVIRKIP 352

- 35 There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 781

- 40 A DNA sequence (GBSx0829) was identified in *S. agalactiae* <SEQ ID 2399> which encodes the amino acid sequence <SEQ ID 2400>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3377 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 50 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



**Example 782**

A DNA sequence (GBSx0830) was identified in *S. agalactiae* <SEQ ID 2401> which encodes the amino acid sequence <SEQ ID 2402>. This protein is predicted to be homology to cl-like repressor. Analysis of this protein sequence reveals the following:

```

5 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.0827 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:AAD44097 GB:AF115103 orf122 gp [Streptococcus thermophilus
 bacteriophage Sfi21]
 Identities = 57/125 (45%), Positives = 77/125 (61%), Gaps = 5/125 (4%)

 Query: 3 MKLDQLCKEFGVLCFSDASDNHSSGFYNPITVLGVGVNLSEQEQKQVALHELQKHNF 62
 M ++L ++FGV LC F +S W GF +P+ +V+ ++ +L ++ +V LHEL H H
20 Sbjct: 1 MNSEKLLAQFGVSLCKFSSSQWTRDGLDFVNRVYINRDLPTERRLAVLLHELGHLEHD 60

 Query: 63 PQYQLFRERCELDANRNMIIHLLKBELEIAEDHTQFNVLVFMKEYLKTIADEAMIKKE 122
 P QY+ RE+ E ANRNMIIH LLK E+ FNY+ FMKEY L TI DE +K E
25 Sbjct: 61 PQYERLRKYEACANRNMIIHELLKN-----ENLGNFYVHFMEKYNLTTCDETFTVINE 115

 Query: 123 YLNLV 127
 YL L+
 Sbjct: 116 YLILI 120

```

30 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 783**

A DNA sequence (GBSx0831) was identified in *S. agalactiae* <SEQ ID 2403> which encodes the amino acid sequence <SEQ ID 2404>. This protein is predicted to be EpsR protein. Analysis of this protein sequence reveals the following:

```

 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.4692 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 The protein has homology with the following sequences in the GENPEPT database.

```

 >GP:AAF12710 GB:AF066865 repressor protein [bacteriophage TFW22]
 Identities = 36/101 (35%), Positives = 62/101 (60%), Gaps = 7/101 (6%)

 Query: 4 LIDRIRELNSKNGMSLNDLEDTLGYSRNSLYSINE--NSRMGKPKETAQYFNVSLDYLLGL 62
 L ++I+EL+++K +S+ +E+ LG++ ++ + N + K K++A+YFNVS+D+LLGL
50 Sbjct: 3 LYEKIKELASQKNVSIQVBEKLGFGANGTITRQMGKKNPQINKVKIDVAKYFNVSVDFLLGL 62

 Query: 63 TUNPRIAS--DTAIIIDQGVVDLREANAHTMLFDGKPLSD 101
 DN R D +D V+ E + FDGKPL ++
55 Sbjct: 63 DNNQRKEFPVDLADFPVDNKNVNDWVS----FDGKPLSD 99

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 784

A DNA sequence (GBSx0832) was identified in *S.agalactiae* <SEQ ID 2405> which encodes the amino acid sequence <SEQ ID 2406>. Analysis of this protein sequence reveals the following:

Possible site: 43  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4079 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 785

A DNA sequence (GBSx0833) was identified in *S.agalactiae* <SEQ ID 2407> which encodes the amino acid sequence <SEQ ID 2408>. Analysis of this protein sequence reveals the following:

Possible site: 52  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2942 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10171> which encodes amino acid sequence <SEQ ID 10172> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 786

A DNA sequence (GBSx0834) was identified in *S.agalactiae* <SEQ ID 2409> which encodes the amino acid sequence <SEQ ID 2410>. This protein is predicted to be a replication initiation protein Rep (RC). Analysis of this protein sequence reveals the following:

Possible site: 54  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3335 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 5 Example 787

A DNA sequence (GBSx0835) was identified in *S.agalactiae* <SEQ ID 2411> which encodes the amino acid sequence <SEQ ID 2412>. This protein is predicted to be antirepressor. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3380 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
15 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA97816 GB:AB044554 antirepressor [Staphylococcus aureus
 prophage phiPV83]
20 Identities = 70/153 (45%), Positives = 93/153 (60%), Gaps = 15/153 (9%)

Query: 3 EIFVPHQGEVRTVTINNEPFWGKDADILGYSKSRNAIALHVDEDDALKQGITNGLGRM 62
 + F F VRTV I NEP-FVGKD+A-ILGY+++ NAI HVD +D L ++ G+
25 Sbjct: 5 QTENFKELFVRTVEIENEPYFVGKDIAELGYARTDINRNVHDSBDKLTQFAS-QQN 63

Query: 63 QETIIINESGLYSIL-SSKLPQVKE---FKRWVTSEVLPQIQGAYVPENLSDR 114
 + IINESGLYSLI SK ++R FKRWVTSEVLP IR+ G Y +N+ ++
30 Sbjct: 64 RNMIIINESGLYSLIFDASKQSKNEKIRSTARKFKRWVTSEVLPFAIRKHGIYATNVIEQ 123

Query: 115 A-----FIALPTQCKKLKEHQLALAQFVDYLK 141
 I + T KK KE L L Q V+ K
30 Sbjct: 124 TLKDPDYIITVLTXYKKEKQNLVLQCCQVEVVK 156

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2413> which encodes the amino acid sequence <SEQ ID 2414>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4609 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 54/142 (38%), Positives = 73/142 (51%), Gaps = 7/142 (4%)

Query: 11 EVRTVTINNEPFWGKDADILGYSKSRNAIALHVDEDDALKQGITNGLGRMQETIINE 70
 EVRT TNN+ +F D IL S R I +++D I D+LGR Q+ INE
50 Sbjct: 13 EVRTATINNGIYFNLNDCQIILELSPNPKYIE-RINKRGVTTSDIISLGRTOQANFINE 71

Query: 71 SGLYSILSSKLPQVKEFKRWVTSEVLPQIQGAYVPENLSDR-----FIALPTQCK 124
 S Y L+ S+ P+ +F WVTSEVLP IR+ GAY+ E ++A I L K
50 Sbjct: 72 SNPYKLVQSRKPFARKFADWVTSEVLPSEIRKHGAYMTGFLRQALTSDFPLIRIANELK 131

Query: 125 KLEKHQLALAQVDYLAKEQPI 146
 + KE L + L E +
55

```

Subjct: 132 EEJERSRQLAEKSIILSVENMV 153

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 5 Example 788

A DNA sequence (GBSx0836) was identified in *S. agalactiae* <SEQ ID 2415> which encodes the amino acid sequence <SEQ ID 2416>. This protein is predicted to be e11. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3281 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA027227 GB:AF009630 e11 [bacteriophage bIL170]  
 Identities = 66/161 (40%), Positives = 93/161 (56%), Gaps = 13/161 (8%)  
 Query: 15 YQVSNLGRVSIQRTVNAKQTRKTGRILKQSL-SSGYAIVTLVSNGLRKSIIRVRLVA 73  
           Y+VSNLG+VR+I                  GRILK + +GY + L N +K++ +HR++A  
 Subjct: 16 YEVSNLGRVNI-----KSGRILKPWIVPVGYLHQLCENNKKNLFLHRIIA 63  
 Query: 74 ERFIPWPINKRTINHIDENKLNKRVNLEWATDKENANHNRTTKSSLRCKPVEQFTLE 133  
           AFI NP K +NHIDENKLN +NLEW T KEN HGR + + K V Q L  
 Subjct: 64 TAFIDNPEEKQVNHIDENKLNLDNLEWCTVKENNIHGTMRKRIAEGHFKKVIQLDLN 123  
 Query: 134 GEFINTFDISKASMCIGISSQRITATAMGHCKQTHGYKWK 174  
           +N F+S+ A +TG+S + I++ G +K +KWR  
 Subjct: 124 DNVLNEFESMVQAEQSTGVSRNLISSOCNGRKSAGRFKWK 164

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 789

A DNA sequence (GBSx0837) was identified in *S. agalactiae* <SEQ ID 2417> which encodes the amino acid sequence <SEQ ID 2418>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2357 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10169> which encodes amino acid sequence <SEQ ID 10170> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 790

A DNA sequence (GBSx0838) was identified in *S. agalactiae* <SEQ ID 2419> which encodes the amino acid sequence <SEQ ID 2420>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -5.47 Transmembrane 21 - 37 (19 - 38)
```

```
----- Final Results -----
 bacterial membrane --- Certainty=0.3187 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 791

A DNA sequence (GBSx0839) was identified in *S. agalactiae* <SEQ ID 2421> which encodes the amino acid sequence <SEQ ID 2422>. This protein is predicted to be DNA polymerase III delta prime subunit (dnaB). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0544 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
AAF98347 AF280763 DNA polymerase III delta prime subunit [Streptococcus pyogenes]
Identities = 284/444 (63%), Positives = 357/444 (79%), Gaps = 4/444 (0%)
```

```
Query: 3 ELKVLPHDQAEQSVLGSIFIPKPKMIEVAEYLPKPNDFYRPAHKILFKAMVSLDRGEAI 62
 EL+V P D+ AEQSVLGSIFI P+K+I V E++ P+DFY+ AHKI+F+AM++L+DR +AI
 Sbjct: 8 ELRVQPDLLARQSVLGSIFISPDKLIAREFISPDFFYKYAHKIFRAMITLSDNRDAI 67
```

```
Query: 63 DIVTIKSTLESTDELGMVGSISYIAIEIVNAVPTSSHAHYAKIVAKQQLRSIIDNLSDS 122
 D TI++ L+ D+L +GG+SYI E+VN+VPTS++AE+YAKIVA+KA LR II L++S
 Sbjct: 68 DATITRTILDDQDLQSIGGLSYIVELVNSVPTSANAEYYAKTVAERKMLRDIARLITES 127
```

```
Query: 123 IGNAYDEIMDIDRIIAKARSLIEVSGASNKSSFRPIHDVLLNHSKIEKRSRNTSQITG 182
 + AYDE + +E+IA ER+LIE++ SN+S FR I DVL N+ +E RS TS +TG
 Sbjct: 128 VNLAYDEILKPEVIAGVERALIILNHSNRSGRKISDVLKVNVEALEARSKQTSNVIG 187
```

```
Query: 183 IRTGFYDFDKLTIGLHEDQLIVLAARPAMGKITALNIAQNVATKSNKAVAFVFSLEMGAE 242
 + TGF D DK+ TGLH DQL++LAARPA+GKTA LNIQNV TK K VA+FSLEMGAE
 Sbjct: 188 LPTGFRDLDKITITGLHEDQLIVLAARPAVAGKTAFLVNLACNVGTQKQKTAIFVFSLEMGAE 247
```

```
Query: 243 SILVERNLGAEGTIIINHIRTGNLTIVHQRLLIYAQQQLARAPIFIDDTYAGVKITDIRARA 302
 SLV+RNL+REG + +H+RTG LT +W + AOG LAEAPI+IDDT G+KIT+IRAR+
 Sbjct: 248 SILVRMLAAGGMVDSHSIRTKQLTDQWNNVITAGALARAPIYIDDTPGIKITIRARS 307
```

-888-

Query: 303 RRLSQETD-GLGLIVIDYQLQLIQSSRSNRCQEVSEISRLQKIIAKELKVPVIALSQLSR 361  
 R+LSQE D GLGLIVIDYQLQLI G++ +NRQEV+S+ISRLQKII+AKELKVPVIALSQLSR  
 Sbjct: 308 RKLSCQVDDGLGLIVIDYQLQLITGTTKPNRCQEVSDISRLQKIIAKELKVPVIALSQLSR 367

5 Query: 362 GVEQRNDKRPIMSDLRBSGSIEQDADIVAFLYRDATYQ---DKKGQPPENDITELIRK 418  
 GVEQR DKRP++SD+RBSGSIEQDADIVAFLYRD YY+ D E E++ E+I KN  
 Sbjct: 368 GVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYTRKCDADAEVENDITIVILEKN 427

10 Query: 419 RHGNLQTVKLYFKKEYTKFSSVEE 442  
 R G GIVKL P KEY KFS+ +  
 Sbjct: 428 RAGARGTVKLMFQKEYNKFSIAQ 451

There is also homology to SEQ ID 2424:

Identities = 284/444 (63%), Positives = 357/444 (79%), Gaps = 4/444 (0%)

15 Query: 3 ELKVLPHDIQAECVSLGSIPIFKPKMIEVAEYLKPNDFTRPAHKILFKAMVSLADRGSAI 62  
 EL+V P D+ AEQSVLGSIFI P+K+I V E++ P+DFT+ AHKI+F+AM++L+DR +AI  
 Sbjct: 11 ELRVQCDLLAEQSVLGSIFISPKLLIAVREFISPDQFYKHAHKIIPRAMITLSDNDCAI 70

20 Query: 63 DIVTIKSTLESTDELGMVGSIYIAEIVNAVFTSSHAHYAKIVAKQAQLRSIINDLSDS 122  
 D TI++ L+ D+L +GG+SYI E+VN+VETS++AE+YAKIVA+KA LR II L++S  
 Sbjct: 71 DATTIRITLDDQDLQSIGSLSYIVELVNSVFTSANAETAYAKIVAEKAMLRDITAIRLES 130

25 Query: 123 IGNAYDEEMDIDEIIAKAERSLIEVSQASNKSSFRPHDVLLENHKSIBERSNNTSQITQ 182  
 + AYDE + +E+IA ER+LIE+++ SN+S FR I DVL N+ +E RS TS +TG  
 Sbjct: 131 VNLAYDEILKPEVIAGVERALILNHSNRSNGSGFRKISDVLKNVYEALEARSQKTSNVITQ 190

30 Query: 183 IETGFPYDPLKITGLHEDQLVLAARPAMGKTALALNTAQWATKSNKAVAVFSLEMGAE 242  
 + TGF D DK+ TGLH DQL++LAAKPA+GKTA LNTAQNV TK K VA+FSLEMGAE  
 Sbjct: 191 LPTGFRDLDKITITGLHPDQLVLAARPAVGKTAFLVNTIAQNGVTQKKTVAIFSLNDAE 250

35 Query: 243 SILVERMLSAEGTTINHHIKGNLTWVNSQRLIYAQQQLAEAPIFIDDTAGVKITDIRARA 302  
 SLV+RML+ABG +H +RGE LT +N + AQQ LAEAPI+IDDT G+KIT+IRAR+  
 Sbjct: 251 SLVDRMLAABGMVDHSLRTGLQTDQWNNVTIAQALAEAPYIDDTGGIKITAIRARS 310

40 Query: 303 RRLSQETD-GLGLIVIDYQLQLIQSSRSNRCQEVSEISRLQKIIAKELKVPVIALSQLSR 361  
 R+LSQE D GLGLIVIDYQLQLI G++ +NRQEV+S+ISRLQKII+AKELKVPVIALSQLSR  
 Sbjct: 311 RKLSCQVDDGLGLIVIDYQLQLITGTTKPNRCQEVSDISRLQKIIAKELKVPVIALSQLSR 370

45 Query: 362 GVEQRNDKRPIMSDLRBSGSIEQDADIVAFLYRDATYQ---DKKGQPPENDITELIRK 418  
 GVEQR DKRP++SD+RBSGSIEQDADIVAFLYRD YY+ D E E++ E+I KN  
 Sbjct: 371 GVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYTRKCDADAEVENDITIVILEKN 430

Query: 419 RHGNLQTVKLYFKKEYTKFSSVEE 442  
 R G GIVKL P KEY KFS+ +  
 Sbjct: 431 RAGARGTVKLMFQKEYNKFSIAQ 454

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 50 Example 792

A DNA sequence (GBSx0840) was identified in *S. agalactiae* <SEQ ID 2425> which encodes the amino acid sequence <SEQ ID 2426>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2146(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

60

A related GBS nucleic acid sequence <SEQ ID 10167> which encodes amino acid sequence <SEQ ID 10168> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 793

A DNA sequence (GBSx0841) was identified in *S.agalactiae* <SEQ ID 2427> which encodes the amino acid sequence <SEQ ID 2428>. Analysis of this protein sequence reveals the following:

10 Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2774(Affirmative) < succ>  
 15 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 794

A DNA sequence (GBSx0842) was identified in *S.agalactiae* <SEQ ID 2429> which encodes the amino acid sequence <SEQ ID 2430>. Analysis of this protein sequence reveals the following:

25 Possible site: 28  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.91 Transmembrane 63 - 79 ( 62 - 79)  
 ----- Final Results -----  
 30 bacterial membrane --- Certainty=0.1765(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8661> which encodes amino acid sequence <SEQ ID 8662> was also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 10  
 MoG: Discrim Score: -11.31  
 GVH: Signal Score (-7.5): -1.86  
 Possible site: 28  
 >>> Seems to have no N-terminal signal sequence  
 40 ALLOW program count: 1 value: -1.91 threshold: 0.0  
 INTEGRAL Likelihood = -1.91 Transmembrane 61 - 77 ( 60 - 77)  
 PERIPHERAL Likelihood = 9.92 19  
 modified ALLOW score: 0.88  
 45 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1765(Affirmative) < succ>  
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-890-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AB18686 GB:U38906 ORF11 [Bacteriophage x1t]  
Identities = 101/249 (40%), Positives = 157/249 (62%), Gaps = 21/249 (8%)

Query: 3 NAGRRMPSKITTETDRFLEMFSSQALVPHLWGRADRGWIDAKTITQRTIGASDDMKL 62  
NAGRRM ++ +T +FL +EL +QALVPHL + ADDAG ++ A + R +GA++D + L  
10 Sbjct: 1 NAGRRMIDKRTITQTKFLRLPLEYQALVPHLMNADDDGVVR-AFPVVMVGAARDSLGL 59

Query: 63 LIAKGFLLIPDGGVV-VIRHMRIMNYIQSDRFQSTLYQSKAQLEYDKSKTASLAKPTGNC 121  
L+ K F+ P + +V I ++ N I+ DR++++ Y AQL ++ ++P N  
10 Sbjct: 60 LVVQGFPIKPIKRMVVIIFIDFKQNTIKDKYKASKY----AQLITNEEPGTMSPERNQ 115

Query: 122 IQNVSMETQVRLSGSLDKSLITTYTVDNMESEDDIPYKRTISYLINEANRNYPIQK 181  
+ K RL K LDK++ +S ++ IPY EI+ YLN+K R++R N++  
10 Sbjct: 116 LGTSDKN----RLDKNRLDKN----NMSGKPDVIFPYSEILEYLNKTKGRSFR-NVEA 165

Query: 182 NKTLIKARNSRGFRLLDDFKHVIDTYVKNWSTGY----EKYLRPETLPGSKVFGYLNQA 236  
NK LIKARNA+G++L+DFK V+D V +WSG + E YL+P+TLF +KF+ YLNQ  
20 Sbjct: 166 NKCLIKARNSGYKLEDFKTVVDNVMNWSGKMFNGVPAENYLPQKTLISNKFDSYLNQV 225

Query: 237 PRIKTTED 245  
PRI+ + I+  
25 Sbjct: 226 PRIQKBN 234

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8662 (GBS344) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 12; MW 30.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 3; MW 59kDa).

The GBS344-GST fusion product was purified (Figure 213, (lane 3; Figure 226, lanes 4-6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 271), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 795

A DNA sequence (GBSx0843) was identified in *S.galactiae* <SEQ ID 2431> which encodes the amino acid sequence <SEQ ID 2432>. Analysis of this protein sequence reveals the following:

Possible site: 47  
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2549 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

50 >GP:RAG31329 GB:AF182207 ORF 272 [Bacteriophage mv4]  
Identities = 70/241 (29%), Positives = 125/241 (51%), Gaps = 30/241 (12%)

Query: 12 VLESTCEVHCQGLWTKVPIKGRLEHKKQCECTKAANIPENKLNQSKINSKLADTYA 71  
VLE+ C HG L +T +G E++ CP+C A+ + + + + + + + S +A  
Sbjct: 16 VLEQKCSKHGLNL-ITTKNHG--BGVTCPCQQAALSVLQSRPDQKAR-QSITARK-- 69



-891-

Query: 72 VFPERDSELVSDKLRAKSLNVE-----IKDEIDQHALNYAKRMQFYRQRTGNAIL 122  
 F +SL + K+ + + +E IK ++ A+ A+ + A++  
 Sbjct: 70 -FRENSLANSKMWCTITFDTPAQPQSAEELIKGVORNAVAFAATKPAVHH-----AVL 121

5 Query: 123 TGRSGVGKSHLTYGLAKPQTEQPKAYBSKVLFISSLVSLPTKIKESFKVDNRY-RQADM 181  
 G G GKSEL A M ++ + K+ + FI++ LF+KIK SF + Y +  
 Sbjct: 122 YGQRGAGKSHL---ANAMMGIHKHPTKMTAPINISRLFSKIKNSFDOPSEYWTKEKA 177

10 Query: 182 ISLLTRVVDYLFLDDLGKESRKQDS--QNNWTHQILYELLNRSNTIINTHLSKKEIKALY 240  
 +E+ VD L +DDLG ES G + + +W ++Y+L+N+ II THLS +E+K +Y  
 Sbjct: 178 LEIMRGVOLLICDDLGESSNGRTGQATKMAQDVIVYDVLNEDRIIITTHLSERELGRVY 236

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 15 vaccines or diagnostics.

### Example 796

A DNA sequence (GBSx0844) was identified in *S. agalactiae* <SEQ ID 2433> which encodes the amino acid sequence <SEQ ID 2434>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

20 Possible site: 47  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1241 (Affirmative) < succ>  
 25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10165> which encodes amino acid sequence <SEQ ID 10166> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC98421 GBI: L29323 methyl transferase [Streptococcus pneumoniae]  
 Identities = 262/474 (55%), Positives = 313/474 (65%), Gaps = 71/474 (14%)

35 Query: 2 MKFLDLFAGIGGFLRMGEAGHECIGPCEINKPARASYKVIHDTSEGEIHLHDITRVD-E 60  
 M+P+DLF+GIGGFLRMGE GHECIGPCRI+KPAR SYK I TGEIE HDI VSD E  
 Sbjct: 1 MRFLDLFSGIGGFLRMGSVGHECIGPCEIDKFAESYKSIQFTEGEIEFDITRVDSE 60

Query: 61 FIRGISVDVIOGGFPOQAFSLAGNRGFPEDTRGTLFFELARFASILEPKYLFLENVGL 120  
 F + G VDVIQGGFPOQAFSLAG R GPEDRGTLFFELAR A ++P+LELENVGL  
 40 Sbjct: 61 PKILAGKVDVIOGGFPOQAFSLAGRRIGFEDTRGTLFFELARAAKIQPLFLLENVGL 120

Query: 121 INHEGATFETIIRTLDELGVNVSMQIFNSKNPGVPQNRBRVFIHLRGEGTRPIPFPE 180  
 LNH+ G TF TI+ TLDELG++VEMQ+ NSK+FGVPQNRBRVFIHL R GTR PFF  
 45 Sbjct: 121 INHDKGRTFTITLTLDELGDFVEMQMSKDFGVPQNRBRVFIHLRGEGTRPIPFPE 180

Query: 181 SSITENVPIHTRKIGVNPFGMGNGGEVYDSEGLSPILTNKGSGVKIAVN----- 231  
 P + +GN+NPS +GN+G+VY SEGL+FTL XGSG KIA+  
 Sbjct: 181 REGQATNPFTLILGLMFLPSKSGMSKGVYSEGLAFTLVKSGKGFKIAIPNTDRLDK 240

50 Query: 232 -----VVGRLEPGKFEMPRVYDPGLAFTIRTKMGQGLE 265  
 VVG LP F+ RYV +GL+PT+ TMQGG  
 Sbjct: 241 RQWRRRPKUNGEPMTLTQDRHIVVGLDFTSKPSTKTRGVYSEGLSPILTMQSGDKI 300

Query: 266 PKLIQRGRGVNQSGHYEISPTVTCNSQNNLLKIKEATKKGYSEAEGDSVNLHSPNSE 325  
 PKI+ + LK++EATKKGY++AE GDS+NL P+S+  
 55 Sbjct: 301 PKILLP-----EPIQLKVRKREATKKGYAQAEGIDGSLNLEKPSQ 339

Query: 326 TRRGVGVKGIANTLLAGEEQGVV--YDLNRRKDIVGTITASCHNGNTTGTGTGINS 383

-892-

```

RRGRVGGKLIANTL T + GVVV Y+ +++ + G L G
Sbjct: 340 RRRGRVGGKLIANTL/TSGQMGVVVA5EGEDRQVQVAGV/LID-----GQFYR 387

Query: 384 FRIRK/TPRECNRLLQGFPDWA7DKASQVNSNSQLYKQAGNSVTNVIAAIAARRL 437
RIR++TP+EC+RLQGFPDWA+ A +V+SNSQLYKQAGNSVTV VIAAIA++L
Sbjct: 388 LRIRRLTPKECFRLQGFPDWA7EAAKRVSN5QLYKQAGNSVTVPVIAAIAAKL 441

```

There is also homology to SEQ ID 2436:

```

Identities = 53/75 (70%), Positives = 62/75 (82%), Gaps = 1/75 (1%)

Query: 2 MKFLDLFAGIGGFRIGMBQAGHECIGFCRINKPARASYKVIHDTEGRIELHDITVSEDP 61
MKFLDLFAGIGGFRIGL+ HECIGFCBI+KPAR SYK I++TEGRIE HDI +V+D+
Sbjct: 4 MKFLDLFAGIGGFRIGLINQCHECIGFCRIDKPARQSYKAIYETEGRIEFDIRQVDQD 63

Query: 62 IRGI-GSDVDVIOGF 75
R + G VD+IOGSF
Sbjct: 64 FRQLGGQVDIIIOGSF 78

```

Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

#### Example 797

A DNA sequence (GBSx0845) was identified in *S. agalactiae* <SEQ ID 2437> which encodes the amino acid sequence <SEQ ID 2438>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2585 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 798

A DNA sequence (GBSx0846) was identified in *S. agalactiae* <SEQ ID 2439> which encodes the amino acid sequence <SEQ ID 2440>. This protein is predicted to be arpR protein. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5070 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AB09197 GB:U24159 orf12 [Bacteriophage HP1]
Identities = 34/69 (49%), Positives = 47/69 (67%), Gaps = 1/69 (1%)

Query: 1 NTKIWTLEKVEQWFDNRNLE-ANPQKQKLEKTEGRLYSGLAKGKSEIIRDSLDNQ 59
M L + +EQW DRNL E + P KQF KL+ER GEL SG+AK K ++I+DS+GD

```

-893-

Sbjct: 1 MADLQQILIKNIEQWAEERNLVEDSTPQKQFIKLMERFQHLCSGVAKNKPDVIKDSIGDCF 60

Query: 60 VVLIGIEQQ 68

VV++ ++Q

Sbjct: 61 VVMVILAKQ 69

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

# Example 799

A DNA sequence (GBSx0847) was identified in *S.agalactiae* <SEQ ID 2441> which encodes the amino acid sequence <SEQ ID 2442>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -5.10 Transmembrane 13 - 29 ( 10 - 36)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3039(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD21919 GB:AF085222 unknown [Streptococcus thermophilus  
 bacteriophage DT1]  
 Identities = 31/67 (46%), Positives = 49/67 (72%), Gaps = 1/67 (1%)  
 Query: 42 HQEADRVIIYVADNAGNEMFGKITDKIEIEGRHTVTAGAYGKFLVTEEQYNEITVGEDIP 101  
 ++ + ++++ ADN E+ GK+T K ++ +T+ GAYGKFLV++EQY+ + VGD+IP  
 Sbjct: 34 NRPVETALVVHKADNF-VELHGKVTGKSMVGKLYTTIDGAYGKFLVSKBQYDSVQVGDIEIP 92  
 Query: 102 DYLGKRG 108  
 YLGKRG  
 Sbjct: 93 SYLGKRG 99

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

# Example 800

A DNA sequence (GBSx0848) was identified in *S.agalactiae* <SEQ ID 2443> which encodes the amino acid sequence <SEQ ID 2444>. This protein is predicted to be gene 17 protein. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5428(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA24397 GB:V01146 gene 1.7 [Bacteriophage T7]  
 Identities = 30/72 (41%), Positives = 40/72 (54%)

-894-

Query: 47 DNNVNPISHYQKYGLESIDVLRNPMTPMLKGFYLGNALKYQLRYRKNGLEDLEKARKN 106  
 + V PSHY +E+I+V+ MT E KG+ GN LKY+LR KK+ L L+K  
 Sbjct: 120 EGVTKPSHYMLPDDIEAIEVLARSMTVEQPKGYCFGNILKYRLRAGKSELAYLEKDLAK 179

5 Query: 107 LEMGLIREMEKEK 118  
 D+ E EK K  
 Sbjct: 180 ADFYKELFEKHK 191

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 801

A DNA sequence (GBSx0849) was identified in *Sagalactiae* <SEQ ID 2445> which encodes the amino acid sequence <SEQ ID 2446>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 28  
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 20 bacterial cytoplasm --- Certainty=0.1375 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 802

A DNA sequence (GBSx0850) was identified in *Sagalactiae* <SEQ ID 2447> which encodes the amino acid sequence <SEQ ID 2448>. Analysis of this protein sequence reveals the following:

- 30 Possible site: 31  
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 35 bacterial cytoplasm --- Certainty=0.0087 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10163> which encodes amino acid sequence <SEQ ID 10164> was also identified.

- 40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF26608 GB:AF145054 ORF9 [Streptococcus thermophilus  
 bacteriophage 7201]  
 Identities = 99/148 (66%), Positives = 116/148 (77%), Gaps = 10/148 (6%)

- 45 Query: 5 MINNVVLIGRLTQDVELRYTPSNIANATFNLAVNRNPKAAGDREADPINCVMRQQAEN 64  
 MINN VL+GRIT+D E +YT SNIA A+P+LAVNRNPK+A G+READPINC+VRQQAEN  
 Sbjct: 1 MINNVVLVIGRLTKDPRFKYTGSTIAVASFSIAVNRNPKDANGERADPINCIVIRQQAEN 60
- Query: 65 LANWTKGKGLIGITGRIGQTRSYBNQOGQRIVYTVTEVVADSFQILEKR----DNSTNQASMD 120  
 LANW KKG LIGITGRIGQTRSYBNQOGQR+YTVSVVA++FQ+LE R + N +  
 Sbjct: 61 LANWAKGKGLIGITGRIGQTRSYBNQOGQRVYTVTVSVVAENFQMLIESRAAREGNANISYSQ 120

Query: 121 DQLP-----PSFGNSQPMDISDQDLFF 142  
 Q+P + N QP+DIS DQDLFF  
 Sbjct: 121 QQVPHFARKNTKYSNKQFLDISDQDLFF 148

There is also homology to SEQ ID 1492.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 803

- 10 A DNA sequence (GBSx0851) was identified in *S.agalactiae* <SEQ ID 2449> which encodes the amino acid sequence <SEQ ID 2450>. This protein is predicted to be puff C4B protein. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1203 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 10161> which encodes amino acid sequence <SEQ ID 10162> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 804

- 30 A DNA sequence (GBSx0852) was identified in *S.agalactiae* <SEQ ID 2451> which encodes the amino acid sequence <SEQ ID 2452>. This protein is predicted to be F5M15.19. Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -2.34 Transmembrane 7 - 23 ( 6 - 23)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1935 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 805**

A DNA sequence (GBSx0853) was identified in *S.agalactiae* <SEQ ID 2453> which encodes the amino acid sequence <SEQ ID 2454>. Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4398 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10159> which encodes amino acid sequence <SEQ ID 10160> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 806**

A DNA sequence (GBSx0855) was identified in *S.agalactiae* <SEQ ID 2455> which encodes the amino acid sequence <SEQ ID 2456>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2992 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 807**

A DNA sequence (GBSx0856) was identified in *S.agalactiae* <SEQ ID 2457> which encodes the amino acid sequence <SEQ ID 2458>. Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4639 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB07758 GB:AP001520 unknown conserved protein [Bacillus halodurans]
Identities = 65/184 (35%), Positives = 102/184 (55%), Gaps = 6/184 (3%)

Query: 1 MNIVEPLRDKDDIQMKDYLSWNEKYMYLFLGLGINTGFRVGDILKLVKDVQGHKIVR 60

```

-897-

```

 M V P R D D I Q A + K L + + Y + L F + G I N T G R + + L L K + K D V
Sbjct: 1 MEYVVPFRDQIQAIKRSLLKKSPDYLLFTIGINTGLRISQLLALKIKDVIDGQKPKD 60

Query: 61 EQKTGKYKSIQMTPLKNEIR---KPVKDKRLHEVLPQGRVGNKNKALSYKTVVWFLKRAA 117
 + + + + + K L + P + + E H L F S + + + + Y + K + A A
Sbjct: 61 YLQLESGEIVYLANDQVKGALQFYAHFIEPQBQH-CLPAS-TNPDQPMTRQHAIRLIKQAA 118

Query: 118 EDLGI-DRVGTHTMRKTPGYHYKKYNVADLMGLFHHGSPAVTLIYICVRQDELDTKMS 176
 + G + D + G T H T + R K T G Y H Y + + + L F H H + P A T L Y I + + + E
Sbjct: 119 LOWGLTDQIGTHTLRKTPGYHAYRQVALSILQQRPHHQITPAQTARYIDIANKSQTIPRI 178

Query: 177 NFEL 180
 N + L
Sbjct: 179 NVNL 182

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 808

- 20 A DNA sequence (GBSx0857) was identified in *S.agalactiae* <SEQ ID 2459> which encodes the amino acid sequence <SEQ ID 2460>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

```

- 25 ----- Final Results -----
- |                         |                                        |
|-------------------------|----------------------------------------|
| bacterial cytoplasm --- | Certainty=0.3582 (Affirmative) < succ> |
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ>   |
| bacterial outside ---   | Certainty=0.0000 (Not Clear) < succ>   |

- 30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 809

- 35 A DNA sequence (GBSx0858) was identified in *S.agalactiae* <SEQ ID 2461> which encodes the amino acid sequence <SEQ ID 2462>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

```

- 40 ----- Final Results -----
- |                         |                                        |
|-------------------------|----------------------------------------|
| bacterial cytoplasm --- | Certainty=0.2732 (Affirmative) < succ> |
| bacterial membrane ---  | Certainty=0.0000 (Not Clear) < succ>   |
| bacterial outside ---   | Certainty=0.0000 (Not Clear) < succ>   |

- 45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 810**

A DNA sequence (GBSx0859) was identified in *S.agalactiae* <SEQ ID 2463> which encodes the amino acid sequence <SEQ ID 2464>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1720 (Affirmative) < succ>
10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 811**

A DNA sequence (GBSx0860) was identified in *S.agalactiae* <SEQ ID 2465> which encodes the amino acid sequence <SEQ ID 2466>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2619 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10157> which encodes amino acid sequence <SEQ ID 10158> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 812**

35 A DNA sequence (GBSx0861) was identified in *S.agalactiae* <SEQ ID 2467> which encodes the amino acid sequence <SEQ ID 2468>. This protein is predicted to be terminase large subunit. Analysis of this protein sequence reveals the following:

```

 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2753 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC27181 GB:AF009630 putative terminase subunit [bacteriophage
 bIL170]

```